A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Editorial 1

ARTICLES

Some Factors Affecting the Quality of Therapeutic Relationships 5
Morris B. Parloff

Schizophrenic Intellectual Impairment: Uniform or Differential? 11
Arnold Binder

Experimental Effects of Ego-Defense Preference on Interpersonal Relations 19
Arthur R. Cohen

The Einstellung Test and Performance in Factual Interviewing 28
Kurt W. Back

Assimilative Projection and Accuracy of Prediction in Interpersonal Perceptions .. 33
Richard M. Lundy

Concept Formation and Impressions of Personality 39
Eugene S. Gollin and Sheldon Rosenberg

"Conceptual Level" vs. "Conceptual Area" Analysis of Object-Sorting Behavior of Schizophrenic and Nonpsychiatric Groups 43
Lawrence S. McGaughran and Lewis J. Moran

Expectations, Fulfillment, and Morale 51
Aaron J. Spector

Continued on inside front cover

Volume 52 JANUARY 1956 Number 1

CONTENTS–*Continued*

The Perception of Social Attitudes 57
Samuel J. Messick

Members' Reactions to Apparent Group Approval of a Counternorm Communication 67
Harold H. Kelley and Christine Lipps Woodruff

Social Sensitivity in the Small Task-Oriented Group 75
J. Richard Suchman

The Measurement of Hostility in Clinical Situations 84
Arnold H. Buss, Ann Durkee, and Marc B. Baer

Perceptual Contact With Reality in Schizophrenia 87
Edward Lovinger

Additional Observations on the Social Perceptions of Authoritarians and Nonauthoritarians 92
Alvin Scodel and Maria Livia Freedman

Physiological Need, Set, and Visual Duration Threshold 96
Janet A. Taylor

The Reinforcement of Cooperation Between Children 100
Nathan H. Azrin and Ogden R. Lindsley

The Effects of an "Anxiety-Reducing" Medication on Group Behavior Under Threat 103
John T. Lanzetta, G. R. Wendt, Peter Langham, and Don Haefner

The Effect of Induced Aggressiveness on Opinion Change 109
Walter Weiss and Bernard J. Fine

The Influence of Reinforcement and Punishment in a Minimal Social Situation 115
Joseph B. Sidowski, L. Benjamin Wyckoff, and Leon Tabory

The Experimental Change of a Major Organizational Variable 120
Nancy C. Morse and Everett Reimer

Correlates of Two Types of Xenophilic Orientation 130
Howard V. Perlmutter

CRITIQUE AND NOTES

The Effects of Manifest Anxiety and IQ on Discrimination 136
Jean S. Kerrick

Taylor Scale Characteristics of Volunteers and Nonvolunteers for Psychological Experiments 138
Phillip Himelstein

A Method for Matching Anonymous Questionnaire Data with Group Discussion Material 139
Elliott McGinnies

The Fakability of Two Measures of Prejudice: The California F Scale and Gough's *Pr* Scale 140
Norman D. Sundberg and Warren D. Bachelis

Volume 52 JANUARY 1956 Number 1

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, Social Science Research Council

Associate Editor: E. J. SHOBEN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHILET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Center for Advanced Study in the Behavioral Sciences

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
New York University

MORTON DEUTSCH
New York University

I. E. FARBER
State University of Iowa

LEON FESTINGER
Center for Advanced Study in the Behavioral Sciences

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

GARDNER LINDZEY
Center for Advanced Study in the Behavioral Sciences

DAVID C. MCCLELLAND
Wesleyan University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D. C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1956 by the American Psychological Association, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis; those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area, this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Social Science Research Council
230 Park Avenue
New York 17, N. Y.

Fifty free offprints are supplied to contributors of articles. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

EDITORIAL

With this fifty-second volume, the Journal comes once more under new editorial direction. On such an occasion, it is traditional and fitting for the new Editor to lift for a moment the cloak of anonymity to make known his conception of his task and the policies that he proposes to follow.

In the nature of the case, these policies will largely be familiar ones. As an elected custodian entrusted with the Journal by the Council of the American Psychological Association, an editor is limited in the degrees of freedom open to him. Any impulse to strike out on uncharted paths must be carefully evaluated in light of his responsibility to the American Psychological Association and to psychology generally. As I see it, these responsibilities require him to work toward the simultaneous furtherance of a fourfold set of interests.

First of these is the tradition of the Journal, which is a long and worthy one and deserves loyalty. It will be just fifty years ago in April that Dr. Morton Prince issued the first number of the *Journal of Abnormal Psychology*, an anniversary that I have plans to observe more appropriately during the course of the year. In the last two decades, Gordon W. Allport in his double term as editor, and J. McV. Hunt most recently, have carried the tradition forward and adapted it to a rapidly developing field, setting a high standard that a new editor can only hope to maintain.

Because of the terms agreed upon when Prince gave the Journal to the American Psychological Association in 1925, I also have a more specific obligation: it was expressly stated in the agreement to which Prince and the representatives of the APA subscribed that the interests of abnormal psychology shall never become secondary. What was abnormal psychology to Prince and his contemporaries, to be sure, has flourished and been transformed in the intervening years. Under this rubric, the Journal that Prince established sought to foster a dynamic psychology. It is a sign of the maturing of psychological science that to do justice to his intent, *abnormal* psychology must at mid-century be interpreted to include contributions from whatever source to the scientific study of the pathology, dynamics, and development of personality. The lessons of Freud and of Janet, exemplified in these pages from the earliest years, have been well learned; the barrier between the normal and the abnormal as subjects of scientific study is virtually dissolved. In the foregoing interpretation of the prior part of the Journal's scope, I of course follow policies made explicit by Dr. Hunt.

More recent tradition also defines the other half of the Journal's broad spectrum: social psychology. Since this very active field in its most inclusive sense embraces major sectors of two disciplines with incursions into still others, selectivity must necessarily be exercised here to keep the Journal in balance. Its focal concern with social psychology will remain the scientific exploration of interpersonal processes, where social psychology tends to converge with the dynamic psychology of personality. Concrete descriptions of social institutions and structures, and analyses of public opinion and mass communication, will continue to be regarded as more appropriate for publication through other channels. No attempt will be made, however, to fix rigid boundaries.

Appropriate specialization of function among the publications of the American Psychological Association also dictates that in both the social and the abnormal and dynamic fields, the emphasis of this Journal must be on articles carrying forward the scientific foundations of understanding rather than the applied techniques of social and clinical practice. While practice may afford the setting and even provide the subject matter for studies directed toward the illumination of underlying principles and processes—studies eminently appropriate to these pages, the *Journal of Consulting Psychology* and the *Journal of Applied Psychology* are the proper channels for more practically oriented contributions.

If an editor finds his bearings in the past and draws sorely needed strength from a vigorous tradition, two of the interests that he serves are very much in the present: authors and readers. The former are becoming so numerous and prolific that it sometimes seems that they must outstrip the latter. A well-edited journal must behave responsibly toward authors, even when on the average about one new manuscript crosses the editor's desk

every calendar day. Although considerably more manuscripts must be rejected than can be published if the Journal is to stay within its allotted space, the element of arbitrariness in editorial decision should be reduced to a minimum, and the reasons for decision made explicit. Correspondence with authors needs to move promptly, and the publication lag to be kept within bounds—preferably under a year. To do justice to the interests of authors—or to make an honorable attempt at justice—in the present expanded state of research in the Journal's field exceeds the capacities of a single volunteer editor.

So as to bring the task within feasible limits, I have been authorized to appoint an Associate Editor to share a considerable amount of the primary editorial responsibility, and am most fortunate that E. J. Shoben, Jr., has agreed to accept this role. A word about how we are dividing the labor is in order. All new manuscripts should be sent to the Editor, as in the past. As Editor, I give all these manuscripts a first reading. Those papers that are primarily concerned with patient groups and with clinical problems and settings—approximately a third of the total intake—I refer to Dr. Shoben, to whom I have delegated full authority to correspond directly with authors concerning acceptance, rejection, and revision. To maintain coordinated editorial direction, we confer personally and frequently in regard both to decisions on particular papers and to general matters of policy in this area of subject matter. Both of us draw on the Journal's Consulting Editors and on other outside consultants when specialized competence and advice is needed. Since the Editor and the Associate Editor both read the more clinical manuscripts, such consultation is less frequently necessary in their instance than with the remaining two-thirds, for which sole responsibility remains with the Editor. The list of Consulting Editors carried on the Journal's masthead reflects this state of affairs; it will doubtless be augmented from time to time.

These Consulting Editors, and the many additional persons who respond to my occasional call for advice, are the truly unsung collaborators for whose essential and necessarily anonymous labors the Editor and his Associate, authors, and readers alike must remain deeply indebted. Prior to assuming the editorial role, I had no conception of the care, thoughtfulness, and painstaking hours that these consultants regularly devote to the evaluation and constructive criticism of the articles referred to them. I continue to be amazed that such good and conscientious counsel is available for me to call upon. The expressions of gratitude that, again to my pleasant surprise, I frequently receive from authors whose research reports have been returned for revision with searching criticism or even rejected, belong in all rights to these individually nameless consultants.

To a large extent the authors' and the readers' interests converge. But readers inevitably suffer in one respect from the growth of the Journal to an enterprise of its present magnitude. It is no longer possible for an editor to give the tender loving care to the literary style of each manuscript that yesterday's readers of a smaller Journal may have come to expect. The burden of observing good standards of readability must rest increasingly on the author. The Editor and his Associate will continue to do what they can to maintain readability, and in this regard as in so many others, they are indebted to the Managing Editor, Lorraine Bouthilet, and her staff in the Central Office of the APA, whose contribution to the production of a clear and stylistically uniform product is a major one.

In reconciling and furthering the interests of the past as represented in the Journal's tradition, and of the present, in its authors and readers, the more timeless interests of psychological science must of course be decisive. It is from this perspective that the conventional criteria for the selection of manuscripts within the Journal's field of subject matter enter the picture: importance of contribution and technical adequacy of workmanship. And of these, it is *importance* that is the crux of the matter and the source of difficulty. No editor, drawing on the very best of specialized advice, can pretend to approach infallibility in foreseeing what will come to be regarded as significant, what trivial or a bypath, in the longer history of his science. Particularly in a field so fluid as that of this Journal, substantial fallibility of judgment is virtually predictable. Yet selection must be made, and it would be an unfortunate abdication of editorial responsibility to rest on the more comfortable criterion of

technical adequacy alone. In practice, it seems to me, an editor can only exercise the best judgment he can muster, attempting to correct for his areas of ignorance and his personal biases of interest and theoretical perspective through frequent and judicious use of consultation. His presumption should and will leave him uncomfortable, but such is his role. Fortunately, scientific communication is a pluralistic enterprise. If the merit of a potentially important piece of research eludes an editor, there are other courts to which its misjudged author can take his plea. The probability of erroneous judgments of importance is, I suspect, part of the necessary cost of responsible editing. At any rate, it is a necessary assumption for the functioning of this Editor.

Considerations of scientific importance are highly relevant to the question of acceptable length of articles, a matter of considerable moment in the case of a journal like this in which an ever-increasing flow of submitted manuscripts presses against the allotted space. I think that the interests of the science, and of author and reader, will be served by the occasional publication of longer and more substantial reports than have become customary. Since one longer piece competes for pages otherwise available to several briefer reports, the significance and merit of such a manuscript must be subject to the closest editorial scrutiny. A single coherent presentation, on the other hand, may often be more economical than the several briefer articles into which it would otherwise be divided. The same standards of conciseness and economy of presentation, it should go without saying, apply to longer and shorter manuscripts alike. For the technically adequate report of research of relatively minor though appreciable consequence, the brief article or note, supplemented by the filing of essential data and procedural description with the American Documentation Institute, seems to me particularly appropriate. If the pressure of submitted manuscripts continues to rise, I expect to rely increasingly on this device to preserve the integrity of the Journal within its present bounds.

Experience with the 231 manuscripts that have come to me in my first eight months of editorial duty indicates that a word may be useful on a frequently recurring problem: When are negative results worth reporting? or, to put it more operationally, When are reports primarily characterized by negative results likely to be acceptable for publication in this Journal? Too often the failure to obtain anticipated results occurs in a context in which the procedures employed are of questionable adequacy and the measures of uncertain relevance or reliability. In such instances, interpretation is inherently so ambiguous that the report is properly judged to fall below the threshold of acceptable significance. Also unacceptable are expositions that purport, contrary to the canons of statistical reasoning, to prove the null hypothesis that specific phenomena are independent of one another. The case is different with studies that obtain negative findings in the course of replicating previously published work in which positive findings have been reported. Provided that replication is exact in relevant particulars—or that the methods as described are ones demonstrated to be relevant and reliable for the purpose—failure to confirm an anticipated relationship can represent a contribution of major importance, reopening questions that had been prematurely closed. Since positive findings in agreement with prior hypotheses tend, other things equal, to support the validity of the procedures employed, it is appropriate that the technical adequacy of studies that come out negatively be more searchingly scrutinized.

With one exception, the contents of the Journal will continue to fall in the same formal categories as before. With the launching of the new *Contemporary Psychology: A Journal of Reviews* concurrently with Volume 52 of this Journal, the Book Review Section that has been under the able direction of Stanley G. Estes since 1948 is brought to a close. The Journal will continue to publish a section of Critique and Notes; indeed, as I suggested above, it may come to rely more heavily on this section than in the recent past. It will also publish Case Reports from time to time. These will usually be evaluated not with respect to their standing as scientific evidence for general propositions (which I believe is generally low) nor yet as contributions to an ideographic human museum (for a variety of reasons, I think, an unfeasible goal); they will be judged, rather, in terms of their heuristic value in calling attention to problematic phenomena, in illustrating important theoretical conceptions,

or in sharpening issues for research. Case reports that meet this criterion should often be valuable in teaching, as they have been in the past.

I should like to express my gratitude to the retiring editor, J. McV. Hunt, who has cooperated most patiently with my attempt to learn from his experience. The articles in this number that bear a date of receipt in January, 1955, are the final product of his Editorship. My responsibility begins with papers bearing a February date. Suggestions and comments on the conduct of the Journal while I am its custodian will be most welcome.

M. BREWSTER SMITH

SOME FACTORS AFFECTING THE QUALITY OF THERAPEUTIC RELATIONSHIPS[1]

MORRIS B. PARLOFF

National Institute of Mental Health

THE interpersonal relationship between the therapist and patient is widely accepted as basic to psychotherapy. It is assumed that the "better" this relationship, the greater the possibility that the patient will benefit from treatment. Despite this, little attention has been devoted to the problem of matching therapist and patient in order to enhance the quality of the relationship. Each school of therapy tends to attribute to itself techniques and skills that presumably enable it to establish uniquely therapeutic relationships (1, pp. 32–47; 2, pp. 280–283; 10, p. 383, p. 389; 11, pp. 350–357; 14, pp. 291–308; 18, pp. 40–45; 21, pp. 115–117). Recent studies of Fiedler (6), Quinn (17), and Heine (13) have raised considerable doubt regarding the validity of such claims. These studies suggest that the "goodness" of a relationship is a function of the "expertness" of the therapist rather than his membership in a particular school of therapy. In comparing relationships established by "experts"[2] and "nonexperts" of the psychoanalytic, nondirective, and Adlerian schools, Fiedler found that experts established "better" relationships than their nonexpert colleagues; moreover, the relationships of experts of one school more closely approximated those of experts of other schools than they resembled relationships created by nonexperts within the same school (8).

These findings suggest that the prediction of the relative "goodness" of the relationship that one or another therapist would establish with a particular patient would be improved by taking into account the variable of the therapist's "expertness." Assuming that expertness is an important factor, one may ask what other variables may influence the quality of the therapeutic relationship established by expert therapists.

Clinical experience, as well as experimental studies such as those of Grob (12) and Fiedler and Senior (9), suggests that the quality of the relationship created between patient and therapist may also be a function of particular aspects of the therapist's personality and his perception of the patient. Would the consideration of such variables enhance the prediction of the quality of the therapeutic relationships that equally expert therapists are able to establish in their treatment of the same patients?

PROBLEM

This is a report of an investigation of the quality of the relationships that each of two expert therapists established with the same patients in group therapy as it related to (*a*) each therapist's personality and (*b*) each therapist's perception of the patients.

Clearly not all aspects of the therapist's personality are equally relevant to the problem of predicting the quality of the therapeutic relationships. That the therapist's capacity to establish ordinary social relationships may be pertinent to this question is suggested by Luborsky, Holt, and Morrow (16), who found that of 247 psychiatrists in training, those who were rated as the better therapists established better relationships with their supervisors and their fellow residents. In a more recent report, Luborsky (15) adds that the therapists who were subsequently rated as most competent were the ones who had also established the better relationships with ward personnel and with the research project staff. It was noted that the ratings of personal liking for the therapist by the research staff were better predictors of the doctor's competence as a thera-

[1] This study represents a portion of a doctoral dissertation completed at Western Reserve University. The research was conducted at The Henry Phipps Psychiatric Clinic of The Johns Hopkins Hospital, Baltimore, Maryland. The author wishes to express his appreciation to Drs. J. D. Frank, E. Ascher, H. Kelman, and D. Rosenthal, and to Messrs. E. Nash and A. R. Stone of the Group Psychotherapy Research Project for their cooperation and suggestions. Special thanks are extended to Dwight W. Miles of Western Reserve University for his patient guidance of this study.

[2] In this study Fiedler's definition was accepted, "An expert is a therapist who has completed the prescribed training of a particular school and who has been a practicing therapist for a minimum of five years and who is considered an expert by other therapists within his school" (6, p. 9).

pist than were the other measures used (15, pp. 3–4). Also, Fiedler, noting that therapists and laymen described their concepts of the "Ideal Therapeutic Relationship" in similar terms, concluded, "... a good therapeutic relationship is very much like any good relationship" (7, p. 244).

It is generally recognized that psychotherapists may have preferences among patients. An expert's "range of acceptance" of patients may be quite broad without his being able to work equally well with all patients. Further, on the basis of his experience, the expert therapist is in a position to evolve a stable concept of his "Ideal Patient," i.e., the hypothetical patient he is most comfortable in treating. His perception of a patient as approximating this "Ideal Patient" is presumably conducive to the establishment of a "good" therapeutic relationship.

In the light of these considerations, the following hypotheses were set up for testing: (*a*) when two expert therapists of similar training and experience but of differing social attitudes treat the same patients, the therapist who establishes the better social relationships creates the better therapeutic relationships; (*b*) when treating the same patient, the therapist who perceives the patient as approximating his "Ideal Patient" concept more closely is the therapist who establishes the better therapeutic relationship.

METHOD

Procedure. Nineteen patients were assigned to three groups of approximately equal size for treatment by group psychotherapy.[3] During the first four weeks of treatment each group was seen for two weeks by one therapist and for the next two weeks by the other therapist (see Table 1). The patients were told that this procedure was being followed to determine the best matching of therapist and group. The therapeutic relationships were judged by three judges. One judge was assigned to each group and attended all group meetings. On the basis of his observations, each judge described the relationships established by the therapist with each patient in the group. These descriptions were in terms of the Fiedler Ideal Therapeutic Relationship Q sample (6), which judges sorted immediately following the second and fourth weeks of treatment. These sortings were then correlated with the Fiedler Ideal Therapeutic Relationship to derive a relationship score for each patient. The higher the correlation, the "better" the relationship was considered to be. The relationships established by each therapist with each of the patients in the three groups were then compared.

TABLE 1

ALTERNATE PAIRING OF TWO THERAPISTS AND THREE GROUPS DURING INITIAL FOUR SESSIONS

Group	Therapist for Each Group During Initial Sessions	
	Therapist during first and second sessions	Therapist during third and fourth sessions
I	Dr. X	Dr. Y
II	Dr. X	Dr. Y
III	Dr. Y	Dr. X

Twelve judges who were acquainted with both therapists made independent evaluations of their social relationships with each therapist. On this basis, one therapist was designated as establishing better social relationships than the other.

The association between quality of social and therapeutic relationships was then investigated. The therapists' perceptions of their patients were derived as follows. Before the study began, each therapist described by means of a 60-item Q sample his Ideal Patient concept. Following the second and fourth weeks of treatment, each therapist described each patient in the same manner. These sortings were then correlated with each therapist's Ideal Patient concept. The details of these procedures are given below.

Professional and social qualities of therapists. Two male psychiatrists, Dr. X and Dr. Y, ages 36 and 42, served as the group therapists. Each was a Diplomate of the American Board of Psychiatry and Neurology, had had 12 years of experience, and enjoyed considerable status. Each had had a personal analysis. Neither therapist claimed adherence to a specific school of therapy, but both were identified with the teachings of Adolf Meyer, Harry Stack Sullivan, Norman Cameron, and Erich Fromm.

To obtain a measure of the quality of the "social relationships" of the two therapists, 12 associates of both were informally interviewed regarding the relative "goodness" of the social relationships that Drs. X and Y had established with them. These judges were five psychiatrists, four psychologists, one social worker, and two stenographers. They included employees, students, and colleagues of lesser academic status and, therefore, might be classed as "subordinates." The omission of "superiors" and "peers" was not considered a

[3] Since Fiedler's Q sample was designed and employed in reference to individual therapy, the possibility that group therapists might strive to establish a relationship different from that of the individual therapist was investigated. Each therapist in this study was asked to describe his concept of the "Ideal Therapeutic Relationship" in a group setting by means of a Fiedler Q sample. The therapists' concepts of the Ideal Therapeutic Relationship correlated .86 and .88 with Fiedler's standard. It was concluded, therefore, that the aims of the group therapist in establishing a therapeutic relationship were similar to those of the individual therapist and that the relationship Q sample was appropriate for use in a group therapy setting.

serious limitation since the patient-therapist relationship is frequently viewed as one of "subordinate" and "superior". Further reports by Luborsky, Holt, and Morrow (16) and Luborsky (15) suggest that the "status" of the individuals judging the social relationships does not affect the predictions of the therapist's competence, i.e., his ability to create therapeutic relationships.

Although the judges unanimously agreed that both men were likeable and highly respected, Dr. Y. was consistently described as more tolerant, respectful, open-minded, self-confident, and supporting of others than was Dr. X. Dr. Y, however, was perceived as more restrained and reserved than Dr. X. The 12 judges unanimously indicated that Dr. Y achieved the better social relationships.

Judgment of therapeutic relationships. The therapeutic relationships established by these two expert therapists were described by three trained judges. Each judge had participated in group therapy observation and research for a minimum of one year. Two were clinical psychologists and the third a psychiatric social worker. Two had completed personal analyses. The judges had similar theoretical orientations based on Sullivan's interpersonal relationship theories. Two of the judges were ignorant of the hypotheses to be tested; the third judge was the writer.

After a practice period, the interjudge reliability was tested during the first two weeks of the study by pairing each judge assigned to a given group with another of the judges. In this manner the relationship developed with each of the 19 patients was independently described by two judges. The mean difference between the judgments of paired raters did not approach statistical significance. The correlation (r) between the relationships described by the assigned judges and the secondary judges was .92, indicating a high degree of reliability between judges.

Selection of patients. Perhaps one of the distinguishing features of the expert therapist is that he knows and respects his own limitations and therefore may attempt to exclude from his practice patients with whom he is not able to establish a therapeutic relationship. In the present investigation, therefore, an attempt was made experimentally to control the selection of patients and thereby to prevent the therapist from selecting "congenial" samples.

Nineteen psychoneurotic patients were selected consecutively from a waiting list at the Out-Patient Department of The Henry Phipps Psychiatric Clinic at the Johns Hopkins Hospital. Patients commonly considered to present special problems in the establishment of a relationship were excluded, i.e., frank psychotics, psychopaths, mental defectives, "organics," alcoholics, and drug addicts. The sample included eight men and 11 women, ranging in age from 20 to 51, the mean being 35.2. Diagnostically, 14 were classified as "psychoneurotic disorders,"[4] four as "personality disorders", and one as a "transient situational personality disorder" (3). The occupational classifications represented in the sample included 10 housewives; four salespersons, agents, or clerks; four semiskilled; and one skilled.

Therapy. Each group met once each week for approximately an hour and a half. The therapy focused principally on the immediate interactions of the patients with each other and with the therapist in order to bring to a patient both awareness of his characteristic behavior patterns and understanding of the needs that this behavior serves in the present. Both therapists regarded transference interpretations as the crucial technique of therapy.

DATA-GATHERING TECHNIQUES

Therapeutic relationship Q sample. Fiedler found that the 75 statements in his Q sample pertinent to therapist-patient relationships were similarly rank-ordered by 10 judges of the psychoanalytic, Adlerian, and nondirective schools. He designated this consistent ranking as the Ideal Therapeutic Relationship concept. Twenty-five of the items concern the therapist's ability to communicate with and understand the patient, 25 deal with the concept of "emotional distance" between the therapist and patient, and 25 refer to the "status" of therapist in relation to the patient (7). In the present research the 75 items were used to describe the observed relationships between therapists and patients, and the resulting arrays were correlated with Fiedler's Ideal.

Perception Q sample. The selection of items for the "perception" Q sample was based on an extension of Bion's theory (4) regarding three basic types of group cultures, representing ways in which group members attempt to enhance their security: (*a*) sex or pairing, (*b*) fight or flight, and (*c*) dependency. These three concepts were expanded by Rosenthal (19) and Stock (20), who suggest that behaviorally these concepts could be either directly expressed in group behavior, or resisted and denied. Six personality dimensions were therefore available: (*a*) pairing, establishing close positive personal relationships; (*b*) counter-pairing, avoiding close personal relationships; (*c*) fight, resisting group members through attack; (*d*) flight, avoiding aggression by techniques of withdrawal; (*e*) dependency, reliance on others for guidance and assistance; (*f*) counterdependency, asserting and maintaining one's own independence by assuming leadership or dominance over the group.

Each of the six personality areas was represented in the final Q sample by 10 statements selected from a large population of items descriptive of typical behavior in groups. Statements were submitted to five judges who were asked to classify each statement into one of the six categories defined above. Only those items on which there was 100 per cent agreement were retained in the final 60-item Q sample.

Before the therapist had met the patients, he described by means of this perception Q sample the hypothetical patient whom he would most prefer to treat. The resulting description was designated as the therapist's Ideal Patient concept. The reliability with which Drs. X and Y could repeat their descriptions seven days after the initial sorting was found to be .89 and .86, respectively.

The Ideal Patient concepts of the two therapists correlated (r) .55. In terms of the relative importance ascribed to the six personality areas, the therapists'

[4] One of these patients was later rediagnosed as having a "psychotic disorder."

Ideals yield a rho correlation of .76, which is not significant for an n of six. The therapists agreed, however, that the three most preferred personality characteristics of an Ideal Patient are fight, pairing, and counterdependence.

After the second and fourth weeks of the investigation (the doctors switched groups at those times), each therapist indicated by means of the perception Q sample his perception of each patient. The sortings were then correlated with the therapist's own Ideal Patient concept.

RESULTS

To test the first hypothesis, which concerned the association between the therapist's social and therapeutic relationships, the following procedure was used. The correlations between each therapist's relationship with each patient and the Fiedler Ideal Therapeutic Relationship concept were converted to Fisher z scores and were then averaged for each therapist in each group. To determine the significance of the differences between the two therapists' mean relationships in each group, the t test for correlated samples was applied. Since Dr. Y, according to 12 "social relationship" judges, created better social relationships than Dr. X, it was predicted that Dr. Y would establish better therapeutic relationships in treating the same patients than Dr. X. The hypothesis is clearly supported, since each of the three judges indicated that Dr. Y established significantly better relationships than Dr. X in each group. In Groups I and III the differences are significant beyond the .02 level of confidence, and in Group II beyond the .01 level[5] (see Table 2).

Table 3 shows that the goodness of the relationships established by the expert therapists varied widely from patient to patient.

To permit an operational test of the second hypothesis, concerned with interdependence of the quality of the therapeutic relationship and the therapist's perceptions of the patient, the following prediction was made: The therapist whose perception of a patient correlates with his Ideal higher than does that of the second therapist establishes the better relationship with that patient. The two therapists were classified according to which established the better relationship with each patient. In addition, the therapists were classified according to which therapist's perception of each patient correlated higher with his Ideal Patient concept. This correlation will henceforth be called Patient-Ideal Congruence. It was found that for 14 patients the therapist who established the better relationship also showed the higher Patient-Ideal Congruence. With only four patients did the therapist who created the better relationship show the lower Patient-Ideal Congruence[6]. Using the sign test, this difference is found to be significant beyond the .05 level of confidence.

The correlation between Patient-Ideal Con-

TABLE 2

MEAN DIFFERENCES IN INITIAL RELATIONSHIPS ESTABLISHED BY EACH THERAPIST WITH EACH GROUP

Group	Mean Relationship (z)		Mean Difference (z)	df	t	p
	Dr. X	Dr. Y				
I	.126	.559	.443	5	3.578	$<.02$
II	.072	.675	.603	5	4.534	$<.01$
III	.403	.633	.230	6	3.333	$<.02$

TABLE 3

VARIABILITY OF CORRELATIONS (r) BETWEEN THE JUDGED THERAPIST-PATIENT INTERACTION AND FIEDLER'S IDEAL THERAPEUTIC RELATIONSHIP CONCEPT, FOR EACH THERAPIST WITH 19 PATIENTS

Variability of Correlations	Therapist	
	Dr. X	Dr. Y
Range	$-.28$ to .63	.19 to .73
Median (z)	.10	.41
Mean (z)	.21	.62
SD (z)	.33	.20

[5] That the differences between the relationships created by the therapists were not limited to the initial period of therapy is suggested by the following incidental data. During the 18th meeting of Groups I and III, Dr. X was ill and was replaced by Dr. Y. The judges of these two groups described the relationships that Dr. Y established with the 10 patients who attended the group meetings. These relationships were then compared with those established with the same 10 patients by Dr. X between the 12th and 20th sessions. The mean difference between the relationships offered by the two therapists to the same 10 patients was again found to be in favor of Dr. Y, this time beyond the .001 level in each group.

[6] Of the 19 patients that Dr. X treated, he failed to recall one of the patients sufficiently to describe him by means of the perception Q sample.

gruence and quality of relationship was compared separately for males and females. In the case of Dr. X, it was found that for the male patients the correlation between quality of relationship and Patient-Ideal Congruence was .78 ($p < .05$), whereas for the female patients, his correlation was $-.18$. In the case of Dr. Y, on the other hand, the correlation for males was .24 and for females was .79 ($p < .01$). This finding suggests that Dr. X described predominantly a male Ideal Patient, while Dr. Y described predominantly a female Ideal Patient. If this inference is true, then the predictive value of the Ideal Patient concept might have been further strengthened if each therapist had been asked to designate two ideal patients: one male and one female.

DISCUSSION

Of the two expert therapists, the one who was judged by mutual associates as integrating the "better" social relationships with them was rated by other judges as establishing the better therapeutic relationships when both therapists alternately treated the same patients during the initial period of group therapy. This finding tends to support Fiedler's observation that a good therapeutic relationship is very similar to any good interpersonal relationship (7, p. 244). Since the measure of goodness of social relationships was defined by the spontaneous statements of individuals acquainted with the two therapists, no assumption is made that the specific qualities that appeared to differentiate the therapists were the only or even the most important dimensions. Of the spontaneously supplied criteria of social relationships (tolerance, respect, open-mindedness, self-confidence, support of others, and emotional restraint), however, all but emotional restraint are generally accepted as important qualities in a therapist. Support for the legitimacy of the criterion of emotional restraint may be found in Luborsky's report that therapists who were found to be the most and least competent of a sample of 247 were similarly differentiated on the ground that the more competent therapists were the more contained and reserved (15).

The relationship between degree of expertness and quality of therapeutic relationships established is far from simple. Equally expert therapists may not establish equally good relationships with a given patient. Moreover, the fact that an expert therapist can establish a good relationship with one patient is no assurance that he will do so with a second patient. The quality of relationship established in this study by each therapist with 19 patients varied markedly from patient to patient. This finding is consistent with Fenichel's observation that, "Any honest analyst will admit that even though he is very thoroughly analyzed, he does better work with certain types of patients than with others" (5, p. 580).

Although it was found that the quality of the therapeutic relationship varied with the degree to which the patient was perceived as approximating the therapist's Ideal Patient concept, the failure to provide for male and female ideal patients may have limited the sensitivity of this instrument.

Inasmuch as these data are limited to the initial relationships created by expert therapists with the same patients, no conclusions are warranted regarding the influence of therapist personality or perceptions upon relationships in later stages of therapy. Since the research was conducted in a group therapy setting with psychoneurotics, it is not known to what extent these findings may apply to other forms of therapy or other categories of patients.

SUMMARY

This investigation attempted to determine the degree to which the quality of the therapeutic relationships established by two equally "expert" therapists with the same patient sample varies in terms of the therapist's personality and his perceptions of the patients. Two expert psychiatrists alternately treated 19 ambulatory psychoneurotics who were randomly assigned to three similar groups for group psychotherapy. The quality of social relationships that each of the therapists established with 12 mutual acquaintances of subordinate status was taken as a relevant measure of the therapist's personality. The therapist's conception of his Ideal Patient and his actual perception of the patients he treated were reported by means of a 60-item Q sample constructed to include Bion's concepts regarding typical group behaviors of patients. The therapist's actual perception of

each patient was then correlated with his Ideal Patient concept.

Three judges, one for each group, observed all treatment sessions. Descriptions of the relationships between each therapist and each patient in a group were made by the judges by sorting the statements of the Fiedler Therapeutic Relationship Q sample. This sorting was correlated with the Ideal Therapeutic Relationship to obtain a measure of relationship.

It was found that the therapist who was able to establish the better social relationships also established the better therapeutic relationships. Moreover, the therapist who perceived a patient as approximating his Ideal Patient concept more closely created the better relationship with that patient.

REFERENCES

1. ADLER, A. *The practice and theory of individual psychology.* London: Kegan Paul, 1929.
2. ALEXANDER, F. *Fundamentals of psychoanalysis.* New York: Norton, 1948.
3. AMERICAN PSYCHIATRIC ASSOCIATION, Committee on Nomenclature and Statistics. *Mental disorders.* Washington, D. C.: Amer. Psychiat. Ass., 1952.
4. BION, W. R. Experiences in groups: V. *Hum. Relat.*, 1950, **3**, 3–14.
5. FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
6. FIEDLER, F. E. A comparative investigation of early therapeutic relationships created by experts and non-experts of the psychoanalytic, nondirective and Adlerian schools. Unpublished doctor's dissertation, Univer. of Chicago, 1949.
7. FIEDLER, F. E. The concept of an ideal therapeutic relationship. *J. consult. Psychol.*, 1950, **14**, 239–245.
8. FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436–445.
9. FIEDLER, F. E., & SENIOR, KATE. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. *J. abnorm. soc. Psychol.*, 1952, **47**, 446–453.
10. FREUD, S. *Collected papers.* Vol. 2. London: Hogarth Press & Institute of Psycho-Analysis, 1948.
11. FREUD, S. *Collected papers.* Vol. 5. London: Hogarth Press & Institute of Psycho-Analysis, 1948.
12. GROB, S. A clinical investigation of the role of the therapist in group psychotherapy. Unpublished doctor's dissertation, Harvard Univer., 1951.
13. HEINE, R. W. An investigation of the relationship between changes and responsible factors as seen by clients following treatment by psychotherapists of the psychoanalytic, Adlerian and nondirective schools. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
14. HORNEY, KAREN. *Neuroses and human growth.* New York: Norton, 1950.
15. LUBORSKY, L. The personality of the psychotherapist. *Menninger Quart.*, 1952, **6**, 1–6.
16. LUBORSKY, L. B., HOLT, R. R., & MORROW, W. R. Interim report of the research project on the selection of medical men for psychiatric training. *Bull. Menninger Clin.*, 1950, **14**, 92–101.
17. QUINN, R. D. Psychotherapists' expressions as an index to the quality of early therapeutic relationships established by representatives of the nondirective, Adlerian and psychoanalytic schools. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
18. ROGERS, C. R. *Client-centered therapy.* New York: Houghton Mifflin, 1951.
19. ROSENTHAL, D. Perception of some personality characteristics in members of a small group. Unpublished doctor's dissertation, Univer. of Chicago, 1952.
20. STOCK, DOROTHY. The relation between the sociometric structure of the group and certain personality characteristic of the individual. Unpublished doctor's dissertation, Univer. of Chicago, 1952.
21. SULLIVAN, H. S. *Conceptions of modern psychiatry.* Washington, D. C.: William Alanson White Psychiatric Foundation, 1946.

Received February 11, 1955.

SCHIZOPHRENIC INTELLECTUAL IMPAIRMENT: UNIFORM OR DIFFERENTIAL?[1]

ARNOLD BINDER

Indiana University

THE purpose of the present study was to test the hypothesis that differential impairment of performance on intellectual tasks is a concomitant of schizophrenia. Although the results of previous investigations have been contradictory and generally confusing with regard to this hypothesis, there has been very general agreement with the restricted specification that schizophrenia is accompanied by impaired performance on tests that require abstract conceptualization. On the basis of the many previous studies supporting this latter more restricted hypothesis, the underlying hypothesis of this study could be stated in the more specific form that schizophrenics show greater impairment in tasks that require abstraction and generalization than they do in tasks that require the reproduction, in a familiar and routine manner, of previously learned material.

Reviews of the scores of studies on the concept of differential intelligence test performance associated with schizophrenia may be found in Wells (28), Harris and Shakow (12), Hunt and Cofer (14), Rabin (19), Rabin and Guertin (20), Hunt (13), Magaret (17), and Rotter (21). Supplementing these reviews are more or less specific critiques of the methods used in the various investigations of the hypothesis; among these are the papers of Garfield (6), Jastak (15), Schofield (22), and Harper (10, 11). The inconsistent results of investigations of differential impairment and the concomitant general uncertainty of the concept have been attributed to the following inadequacies of experimental design (among others):

1. A very large proportion of studies have used tests of general intelligence as experimental instruments and, as Cronbach (4) has most clearly pointed out, these are markedly inadequate for differential evaluation. Since such scales as the Stanford-Binet and the Wechsler-Bellevue were designed to provide global measures of performance, their items or subtests have a high percentage of commonality in order to produce a meaningful total score. (In his factor analysis of Stanford-Binet items, for example, McNemar [18] found a heavy *g* saturation.) Furthermore, the relatively large numbers of different kinds of items in these scales has made for shorter subtests with relatively low reliabilities. (See 5, p. 176, and 18, p. 101.) But this state of affairs is just the opposite of that desirable from the viewpoint of obtaining reliable difference patterns of subtest scores. For purposes of differential evaluation one seeks highly reliable subtests which have very little in common so as to maximize differences in scores which reflect true differences and not merely errors of measurement. As Cronbach has summarized the difficulty, "It is not possible to eat one's cake and have it; the correlation among subtests which makes for good measurement of the common factor interferes with [differential] diagnosis" (4, p. 153).

2. Those investigations not using tests of general intelligence have used either instruments based on some variety of the Babcock (1, 2) method or tests of the form originally propounded by Gelb and Goldstein (7), Goldstein (8), and Vigotsky (26). Hunt and Cofer (14) have objected to the use of Babcock-type tests on the grounds that assumptions of constancy of vocabulary performance despite the disorder and of the adequacy of vocabulary score as an estimate of intellectual ability are not valid. One need only glance at the monograph of Goldstein and Sheerer (9) to justify disposing of the methods of Gelb, Goldstein, and Vigotsky for evaluating the hypothesis of differential impairment. Since these tests were designed primarily to determine the procedures used by individual patients to solve problems and only incidentally the accuracy of solution,

[1] From the Veterans Administration Hospital, Palo Alto, California. The author is grateful to the Research Committee of the hospital, generally, and to Dr. Paul McReynolds, particularly, for their invaluable aid in the support of this study. He is also grateful to the members of his thesis committee at Stanford University, Professors Kenneth Little, Quinn McNemar, and Joseph Luft, for their help on the dissertation upon which this paper is based.

the investigators using them have depended almost exclusively upon qualitative analysis. Moreover, these tests have undergone no formal standardization, there are no available data on normal controls, and there has been almost no interest in questions of reliability, validity, and test intercorrelation. In addition to this, as Cameron (3) has pointed out, different investigators have interpreted the very same type of test performance in divergent ways; to some a particular procedure of solution has reflected the loss of the "abstract attitude" while to others it has indicated the very presence of the ability to abstract and generalize.

3. A surprisingly large number of investigators have failed to control such factors as chronological age, sex, education, and intelligence quotient. Such comparatively homogeneous control groups as nurses, highway patrolmen, and attendants have been used with schizophrenic experimentals of markedly heterogeneous social and educational background. Frequently the investigators have not even mentioned the age, sex, education and/or IQ of their experimental and/or control *S*s, but more often they have determined and reported them, only to ignore obvious differences between their groups in the experimental analysis.

Garfield (6) and Jastak (15) have demonstrated the importance of controlling the variables age, education, IQ, and sex in investigations of differential impairment. Garfield emphasized that these factors had been neglected for the most part in previous studies on diagnostic patterns.

4. Many investigators have been content merely to present their results in tabular or graphical form without any attempt to subject the data to statistical test. (Of course those studies involving tests which allow only qualitative evaluation have been quite restricted as to statistical analysis.) The interpretations coming from such studies have been particularly inconclusive. As Harper (10, 11) has pointed out, moreover, many of the statistical tests which have been used have been quite inefficient.

On the basis of these procedural deficiencies in the previous investigations of the concept of differential intellectual performance associated with schizophrenia, it seems reasonable to conclude that the scientific status of the concept is uncertain. This study represented an attempt to avoid the conceptual and experimental errors made in previous studies and to test rigorously this hypothesis of differential impairment.

METHOD

Tests. The Science Research Associates (SRA) Tests of Primary Mental Abilities (Intermediate) were selected as instruments of investigation because of their relatively low intercorrelations and high reliabilities (see 25, p. 13–14). These tests were developed by the Thurstones (24, p. 1) to measure some of the "relatively *independent*, *describable*, and *testable* abilities" which had been isolated previously in factor analytic investigations. Although Thurstone has been able to isolate and identify eight of these so-called "primary mental abilities," the SRA form measures only five of them. These are Verbal-meaning (*V*), Space (*S*), Reasoning (*R*), Number (*N*), and Word-fluency (*W*).

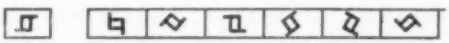
The following is a brief description of the specific characteristics of each of these tests and of the procedures involved in their administration:

a. The Verbal-meaning test is composed of 50 items similar to the following one:

SAFE A. Secure B. Loyal C. Passive D. Young.

The task of the subject is to mark the word which means the same as the first word. A practice period involving five items and about three minutes precedes the test proper. The latter takes four minutes.

b. The Space test is made up of 20 items like the one that follows:



The task of the subject is to mark the figures on the right which are identical to the figure on the left (though rotated), but not to mark mirror images of the left-hand figure. The practice period preceding this test involves six items and about six minutes. The maximum time allowed is five minutes.

c. Reasoning is measured by a test with 30 of the following types of items:

r s r t r u r v r w r x

On each item the letters form a particular kind of logical sequence which the subject must determine. After so doing he must mark the letter that should come next in the series. The total time allowed is six minutes, while about six minutes is used for 10 practice items.

d. Number ability is measured by a series of 60 items similar to the following one:

99
77
82
68

316

The task of the subject is to add each column of figures in order to determine whether the indicated

answer is right or wrong, and to mark the answer accordingly. The time allowed is six minutes. There are five practice items, taking about three minutes, which precede the test proper.

e. In the Word-fluency test, the subject's task is to write as many words beginning with the letter S as he can think of during a five-minute period. The letter M and the letter T are used as alternates in order to minimize the effects of such extraneous factors as previous information and practice. According to Thurstone and Thurstone (25), these substitutions do not affect the norms.

Subjects. The schizophrenic (or experimental) group consisted of a sample of 120 male, U. S. born patients in a Veterans Administration neuropsychiatric hospital. All of the experimentals except one were Caucasian, the one exception being of Japanese extraction.

The experimentals were selected from four wards (two admission and two open) during two (for each ward) separate evaluation periods. Each evaluation consisted of going through the ward folders of all patients, picking out those who had been diagnosed schizophrenic in psychiatric conferences, and then consulting the ward physician for confirmation that the patients actually diagnosed schizophrenic fit the criterion definition adopted for this study. Those patients who met the specifications during any one of the evaluations for the various wards were given appointments for testing. About 10 per cent of these either failed to keep their appointments or left the examination room before the testing was started. Only two patients failed to complete the tests after they answered one or more questions.

Great care was taken to insure the comprehension of instructions and maximum motivation by the schizophrenics. Competence on practice items was a criterion for administering a test proper, testing session groups were restricted to a maximum of four individuals (many consisted of only two subjects and the modal number was three), and the examiner during each session maintained a participating relationship with the subjects.

The normal (or control) sample consisted of 16 patients from a Veterans Administration general medical and surgical hospital and 91 patients from a Veterans Administration tuberculosis hospital. They were male, normal from a psychiatric viewpoint, and native-born American. Of the total of 107 controls, 105 were Caucasian and two were of Japanese extraction.

The tubercular controls were patients who had been hospitalized and treated for pulmonary tuberculosis, but who were, at the time of testing, negative on tubercular salivary and gastric tests. They were on wards which were preparatory stations for their eventual trips home. The other 16 controls were patients being treated for such general medical problems as fractured limbs, kidney disorders, and heart disturbances.

About 5 per cent of the potential control subjects failed to keep their appointments or left the examining room after receiving preliminary instructions. However, the controls who started the testing completed it. Similar precautions to those used to maximize the comprehension and motivation of the experimentals were observed for the controls.

The entire group of normals, like the schizophrenics, were involved in forced hospitalization; they could not leave their respective hospitals, in which they had been resident for extended periods, without medical approval. This seemed an important variable to control and was one of the strong incentives for choosing these particular controls.

The other incentives stemmed from the desirability of getting normal controls who were comparable to the schizophrenics on the variables education, sex, age, and IQ. The degree of adequacy to which this was attained will be discussed below.

Procedure. Since many previous investigators have demonstrated differential intellectual performance between old and young subjects, comparing the entire groups of Experimentals and Controls would not be a satisfactory statistical procedure because of the resulting spuriously high within-groups or error heterogeneity. Consequently, the median age for the entire sample was found, and those Experimentals who were younger than the median age were compared only with the Controls below the median, and those Experimentals who were older than the median age were compared with the older Controls. For the purposes of these comparisons there were 114 *Ss* at or below the age of 39 (55 Experimentals and 59 Controls) and 113 *Ss* at or above the age of 40 (65 Experimentals and 48 Controls).

Table 1 presents the means and standard deviations for both groups of Experimentals and Controls on the variables age and total number of years of education. The statistically insignificant differences between the older Experimentals and older Controls, on the one hand, and younger Experimentals and younger Controls, on the other, makes it reasonable to assume that the samples represent populations equivalent on the parameter values.

There was no way of determining with certainty whether or not there was any difference between the mean IQ's of the prepsychotic Experimental and the Control comparison groups since previous intelligence test results were not available. But the similarity in the composition of the schizophrenic and normal groups in all evaluated variables other than diagnosis makes an assumption in regard to the equivalence of premorbid IQ seem quite justified. The very close agreement between schizophrenic and Control comparison samples on total number of years of education (mean and variance) is particularly relevant.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF TWO YOUNGER AND TWO OLDER GROUPS FOR AGE AND YEARS OF EDUCATION

Statistic	Age				Years of Education			
	Experimentals		Controls		Experimentals		Controls	
	Older	Younger	Older	Younger	Older	Younger	Older	Younger
Mean	53.6	29.2	52.0	27.8	9.9	11.3	9.2	11.2
SD	7.9	5.6	8.0	5.2	3.2	2.2	3.4	2.4
N	65	55	48	59	65	55	48	59

When considered in terms of its implications for the SRA Tests of PMA, the working hypothesis of this investigation could be stated in the following specific form: Schizophrenics show greater impairment in their performance on the Reasoning (*R*) and Space (*S*) tests than they do on the Verbal-meaning (*V*), Number (*N*), and Word-fluency (*W*) tests.

RESULTS

The data (scores on the five tests by each of the various *S*s) were analyzed by means of the analysis of covariance to provide a statistical test of the above hypothesis. The purpose of the test was to determine if the differences between the Experimental and the Control groups on the *R* and *S* tests were significant when allowance was made for the differences accountable on the basis of the differences between the groups on the *V*, *N*, and *W* tests. Since the experimental variables were correlated with three extraneous variables and it was necessary to make the groups comparable with respect to all of these variables simultaneously, the multiple-regression analysis of covariance as described in Snedecor (23) and other statistics texts was adopted. With this sort of analysis, significant differences found between Experimentals and Controls on *R* and *S* beyond those accountable for and predictable on the basis of the variation between the two groups on *V*, *N*, and *W* would support the hypothesis of this investigation. The schizophrenics would show more impairment on tasks that require abstraction and generalization than on tasks that require the reproduction of previously learned patterns of behavior and content associations in a familiar and routine manner.

Four tests of significance were made in evaluating the working hypothesis: the difference between the Experimentals at or below the age of 39 and the Controls at or below the age of 39 on *R* when corrected for the difference accountable on the basis of *V*, *N*, and *W* differ-

TABLE 2

VARIANCE-COVARIANCE MATRIX OF TEST SCORES FOR EACH OF THE FOUR COMPARISON GROUPS

Test	*V*	*S*	*R*	*N*	*W*
V	139.0 92.5	122.5 134.1	44.0 53.1	62.2 44.8	60.8 77.2	73.3 93.9
S	52.8 56.7	97.7 69.7	129.6 109.9	38.7 38.0	38.7 56.6	34.2 67.6
R	30.2 29.1	17.6 24.9	15.0 17.7	55.9 32.5	33.9 39.7	43.4 46.9
N	79.1 58.9	45.6 45.8	23.9 31.4	95.6 100.0	102.6 102.9	72.9 66.6
W	62.5 67.3	39.7 47.8	10.0 35.2	42.4 76.6	159.7 146.6	159.6 158.8

Note. Each number above represents the value $r_{ij}\,\sigma_i\,\sigma_j (i, j = V,S,R,N,W)$ for one of the comparison groups. These four groups are: (*a*) Controls at or below the age of 39, (*b*) Experimentals at or below the age of 39, (*c*) Controls at or above the age of 40, and (*d*) Experimentals at or above the age of 40. The group *a* and *b* variances and covariances are entered above the diagonal in the table, those of group *a* above those of group *b* within each entry cell. The group *c* and group *d* variances and covariances are entered below the diagonal, those of *c* above those of *d*.

ences between the two groups; the difference on R not accountable for on the basis of the variation on V, N, and W between Experimentals and Controls at or above the age of 40; and similarly for the mean differences in S between the younger Experimentals and younger Controls, on the one hand, and the older Experimentals and older Controls, on the other. The variance-covariance matrix for all tests and for each of the four comparison groups may be found in Table 2. Each entry in the table indicates an $r_{ij}\,\sigma_i\,\sigma_j$ value, where $i, j = V, S, R, N, W$. The numbers above the diagonal are the obtained variances and covariances for the two younger groups, the younger Controls above and the younger Experimentals below; and the numbers below the diagonal are the obtained variances and covariances for the two older groups, the older Controls above and the older Experimentals below.

A preliminary comparison was made between the younger Experimentals and younger Controls and between the older Experimentals and older Controls on their respective performances on the five tests, without regard to differential evaluation. Tables 3 and 4 present the results of this comparison.

TABLE 3

COMPARISON OF YOUNGER EXPERIMENTALS AND YOUNGER CONTROLS ON THE FIVE PRIMARY MENTAL ABILITIES TESTS

Tests	Mean of Experimentals	Mean of Controls	Diff. Means	*SE* Diff. Means	*CR****
Verbal-meaning	20.0	30.2	10.2	2.1	4.8
Number	12.8	23.0	10.2	1.9	5.4
Word-fluency	27.4	38.9	11.5	2.4	4.9
Space	11.4	21.5	10.1	2.0	4.9
Reasoning	7.4	13.2	5.8	1.2	4.7

*** $p < .0001$ for all tests.

TABLE 4

COMPARISON OF OLDER EXPERIMENTALS AND OLDER CONTROLS ON THE FIVE PRIMARY MENTAL ABILITIES TESTS

Test	Mean of Experimentals	Mean of Controls	Diff. Means	*SE* Diff. Means	*CR*
Verbal-meaning	16.2	21.3	5.1	2.1	2.5*
Number	13.6	17.6	4.0	1.9	2.1*
Word-fluency	24.3	30.2	5.9	2.4	2.5*
Space	8.4	13.3	4.9	1.8	2.8**
Reasoning	5.2	5.6	.4	.8	.5

* $p < .05$.
** $p < .01$.

TABLE 5

COMPARISON OF YOUNGER EXPERIMENTALS AND YOUNGER CONTROLS ON REASONING (R) AND SPACE (S) BY THE ANALYSIS OF COVARIANCE METHOD

	Source of Variation	Variable df	Variable Sum of Squares	Variable Mean Square	R^2	Errors of Estimate df	Errors of Estimate Sum of Squares	Errors of Estimate Mean Square
Reasoning	Total	113	6063		.611	110	2358	
	Within	112	5087	45.4	.538	109	2350	21.6
	Between	1	976	976		1	8	8

$$F = \frac{8}{21.6} = .37;\ p > .05$$

Space	Total	113	16579		.331	110	11091	
	Within	112	13688	122.2	.220	109	10677	98.0
	Between	1	2891	2891		1	414	414

$$F = \frac{414}{98.0} = 4.23;\ p < .05$$

TABLE 6

COMPARISON OF OLDER EXPERIMENTALS AND OLDER CONTROLS ON REASONING (R) AND SPACE (S) BY THE ANALYSIS OF COVARIANCE METHOD

	Source of Variation	Variable df	Variable Sum of Squares	Variable Mean Square	R^2	Errors of Estimate df	Errors of Estimate Sum of Squares	Errors of Estimate Mean Square
Reasoning	Total	112	1878		.570	109	808	
	Within	111	1873	16.9	.569	108	807	7.5
	Between	1	5	5		1	1	1

$$F = \frac{1}{7.5} = .13;\ p > .05$$

Space	Total	112	9872		.402	109	5903	
	Within	111	9218	83.0	.371	108	5798	55.2
	Between	1	654	654		1	105	105

$$F = \frac{105}{55.2} = 1.90;\ p > .05$$

The differences in means between the Experimentals at or below the age of 39 and the Controls at or below the age of 39 on all test scores are significant beyond the .0001 level. In the case of the older groups (at or above the age of 40), however, three differences be-

tween means are significant at the .05 level (V, N, W), one difference is significant at the .01 level (S), and one difference is statistically insignificant (R).

Tables 5 and 6 contain summaries of the analysis of covariance procedures involved in comparing the Experimentals and Controls in the two respective age groups on R and S. Three of the four final variance ratios led to F statistics not significant at even the .05 level, and the remaining F is just barely significant at the .05 level.

In order to investigate the effects of aging in normals upon intellectual performance decline by the covariance analysis method adopted in this study for testing the hypothesis of differential impairment, the Controls at or below the age of 39 were compared with the Controls at or above the age of 40. Table 7 contains a summary of the statistics involved in this comparison. Since these two groups differed by a little over two years in average number of years education, it was necessary to equate them with respect to this variable before making the final F test. This was done by treating years of education as another extraneous variable and dealing with it by multiple regression analysis of covariance exactly as the variables V, N, and W. This gave four prediction regression variables instead of three. The R-score means for these two groups are significantly different at the .001 level and the S-score means are significantly different at the .01 level, when the groups are made comparable on V, N, W, and years of education.

TABLE 7

COMPARISON OF OLDER CONTROLS AND YOUNGER CONTROLS ON REASONING (R) AND SPACE (S) BY THE ANALYSIS OF COVARIANCE METHOD

	Source of Variation	Variable df	Variable Sum of Squares	Variable Mean Square	R^2	Errors of Estimate df	Errors of Estimate Sum of Squares	Errors of Estimate Mean Square
Reasoning	Total	106	5548		.553	102	2480	
	Within	105	4023	38.3	.479	101	2096	20.8
	Between	1	1525	1525		1	384	384
	$F = \frac{384}{20.8} = 18.46;\ p < .001$							
Space	Total	106	14108		.245	102	10652	
	Within	105	12332	117.5	.193	101	9952	98.5
	Between	1	1776	1776		1	700	700
	$F = \frac{700}{98.5} = 7.10;\ p < .01$							

TABLE 8

COMPARISON OF OLDER EXPERIMENTALS AND YOUNGER EXPERIMENTALS ON REASONING (R) AND SPACE (S) BY THE ANALYSIS-OF-COVARIANCE METHOD

	Source of Variation	Variable df	Variable Sum of Squares	Variable Mean Square	R^2	Errors of Estimate df	Errors of Estimate Sum of Squares	Errors of Estimate Mean Square
Reasoning	Total	119	3078		.612	116	1194	
	Within	118	2939	24.9	.594	115	1193	10.4
	Between	1	139	139		1	1	1
	$F = \frac{1}{10.4} = .10;\ p > .05$							
Space	Total	119	10838		.358	116	6958	
	Within	118	10574	89.6	.344	115	6936	60.3
	Between	1	264	264		1	22	22
	$F = \frac{22}{60.3} = .36;\ p > .05$							

The last comparison on differential decline in R and S made was that between the younger Experimentals and the older Experimentals. Table 8 summarizes the statistics involved in this comparison. Neither the F derived from the between-groups variation on R nor the F derived from the between-groups variation on S reaches the .05 significance level. Again it was necessary to make the groups comparable with respect to number of years of education (by the method discussed above), since the means for the groups differed significantly. However, N did not enter into the regression equation because the older Experimentals did slightly (although not significantly) better than the younger Experimentals on this variable.

DISCUSSION AND CONCLUSIONS

The findings relative to the differences between schizophrenics and normals on each of the five variables, when no consideration was given to differential impairment, are in general accord with the results of previous investigations of over-all schizophrenic intellectual impairment. (Hunt and Cofer [14], on the basis of their survey of the literature, have estimated schizophrenic intellectual deficit at 20 months of mental age.)

The primary hypothesis of this study, that

differential intellectual impairment is associated with schizophrenia, was not supported by the data as analyzed by the covariance method. (If one makes four independent tests of statistical significance and the null hypothesis is in fact true in all four cases, the probability of rejecting the null hypothesis at the .05 level in *at least* one of the tests can be shown by means of the binomial distribution to be equal to .186.)

In the case of the tests of the hypothesis of differential impairment as a result of the process of aging, the null hypothesis was rejected in the comparison of older and younger normals and accepted in the comparison involving older and younger schizophrenics. This would seem to indicate that there is some factor concomitant with schizophrenia that causes great enough over-all impairment of intellectual functioning to mask any differential effects of aging upon the various functions. A similar masking effect may have occurred in the comparisons involving schizophrenics and normals; the gross over-all impairment of the schizophrenics could have produced enough variation between normal and schizophrenic over-all performance to leave little room for further variation between them on the *R* and *S* variables. Certainly the data make it seem reasonable to say that with certain tests (involving a much wider range of possible performance among other things) and with a large enough sample, one might get support for the hypothesis of differential impairment in schizophrenia (especially in a comparison involving younger individuals). The utility of a concept supported under these conditions would of course be very questionable. Since this study involved a series of tests much more suited for the evaluation of differential impairment than those in common usage and sample sizes of respectable numbers, we must conclude that differential impairment is not a useful concept when used in reference to the intellectual performance of patients having the constellation of characteristics subsumed under the name schizophrenia.

The statistical findings of this investigation, as supported by the examiner's observations that, despite all efforts, the schizophrenics as a group did seem somewhat less responsive to the testing demands than the normals, are consistent with the explanations of schizophrenic test performance offered by Hunt and Cofer (14) and by Cameron (3). Hunt and Cofer (p. 1022) explain the differences between normals and schizophrenics on intellectual tasks as follows: "The deficit in the 'functional' psychoses, and particularly in schizophrenia, we conceive as an extinction of standards of performance and of thought skills that have been socially rewarded." Similarly, Cameron has argued vigorously against explanations of the inferior performance of schizophrenics on intelligence tests (whether general or specific) on the basis of a loss of fundamental capacities or a process analogous to organic deterioration. Instead, like Hunt and Cofer, after surveying extensive literature on the subject, he hypothesized that schizophrenics do less well because of diminished social responsiveness and interest in the rewards and expectations of society.

Perhaps one last point should be made. Since all of these tests of primary mental abilities are restricted in terms of time allowed (although they are not strictly speed tests), it might be argued that this accounts for the differences in the performances of schizophrenics and normals and masks the effects of differential patterning. On the basis of the results which she obtained from her administration of the Wechsler-Bellevue to psychotics and normals, however, Magaret (16) concluded that timing and its implications in terms of speed of response were of no such importance. In her words (16, p. 75), "It seems reasonably clear, therefore, that the fundamental explanation of the differences in intellectual functioning among these groups [schizophrenic, paretic, and normal] does not lie in simple speed of response. The psychotics experience difficulty with certain of the timed tests, but not with others, and with certain of the untimed tests, which strongly suggests some more basic explanation of the difference."

SUMMARY

The purpose of this study was to overcome the experimental and statistical inadequacies of previous investigations of the hypothesis that differential impairment of intellectual test performance is a concomitant of schizophrenia. The SRA Tests of Primary Mental Abilities were selected as the instruments of this investigation because they measured the relatively independent abilities postulated in a group factor theory of intelligence.

The 120 schizophrenics and 107 normal controls used in the study were all male, non-Negro, native-born American, and without language difficulty. Both Experimentals and Controls were patients who were required to remain hospitalized at the time of testing, the latter for general medical difficulties.

The 65 schizophrenics in the sample who were at or above the age of 40 at the time of testing were compared with the 48 normals who were at or above the age of 40, and the 55 schizophrenics and 59 normals at or below the age of 39 were similarly compared. There were no significant differences between the schizophrenics and normals (both older and younger groups) in mean and variance of years of education or mean and variance of years of age. The analysis of covariance was used to test the specific hypothesis of differential impairment.

The results of this investigation do not support the hypothesis of differential impairment of schizophrenic test performance. However, severe over-all impairment of schizophrenic intellectual performance and differential impairment of performance on certain kinds of intellectual tasks as a result of aging in normals were clearly shown.

REFERENCES

1. BABCOCK, HARRIET. An experiment in the measurement of mental deterioration. *Arch. Psychol.*, N. Y., 1930, **18**, No. 117.
2. BABCOCK, HARRIET. *Dementia praecox, a psychological study*. Lancaster, Pa.: Science Press, 1933.
3. CAMERON, N. The functional psychoses. In J. McV. Hunt, (Ed.), *Personality and the behavior disorders*, Vol. II, New York: Ronald, 1944.
4. CRONBACH, L. J. *Essentials of psychological testing*. New York: Harper, 1949.
5. DERNER, G. F., ADORN, M., & CANTER, A. H. The reliability of the Wechsler-Bellevue subtests and scales. *J. consult. Psychol.*, 1950, **14**, 172–179.
6. GARFIELD, S. L. An evaluation of Wechsler-Bellevue patterns in schizophrenia. *J. consult. Psychol.*, 1949, **13**, 279–287.
7. GELB, A., & GOLDSTEIN, K. Ueber Farbennamenamnesie. *Psychol. Forsch.*, 1925, **6**, 127–186.
8. GOLDSTEIN, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **96**, 575–588.
9. GOLDSTEIN, K., & SHEERER, M. Abstract and concrete behavior; an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
10. HARPER, A. E. Discrimination of the types of schizophrenia by the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1950, **14**, 290–296.
11. HARPER, A. E. Discrimination between matched schizophrenics and normals by the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1950, **14**, 351–357.
12. HARRIS, A. J., & SHAKOW, D. The clinical significance of numerical measures of scatter on the Stanford-Binet. *Psychol. Bull.*, 1937, **34**, 134–150.
13. HUNT, H. F. Clinical methods: psychodiagnostics. *Ann. Rev. Psychol.* 1950, **1**, 207–220.
14. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt, (Ed.), *Personality and the behavior disorders*, Vol. II, New York: Ronald, 1944.
15. JASTAK, J. Problems of psychometric scatter analysis. *Psychol. Bull.*, 1949, **46**, 177–197.
16. MAGARET, ANN. Differences in intellectual functioning among normal, paretic, and schizophrenic adults. Unpublished doctor's dissertation, Stanford Univer., 1941.
17. MAGARET, ANN. Clinical methods: psychodiagnostics. *Ann. Rev. Psychol.*, 1952, **3**, 282–320.
18. MCNEMAR, Q. *The revision of the Stanford-Binet Scale*. New York: Houghton Mifflin, 1942.
19. RABIN, A. I. The use of the Wechsler-Bellevue Scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410–422.
20. RABIN, A. I., & GUERTIN, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–248.
21. ROTTER, J. B. Clinical methods: psychodiagnostics. *Ann. Rev. Psychol.* 1953, **4**, 295–316.
22. SCHOFIELD, W. Critique of scatter and profile analysis of psychometric data. *J. clin. Psychol.*, 1952, **8**, 16–22.
23. SNEDECOR, G. W. *Statistical methods applied to experiments in agriculture and biology*, (4th Ed.) Ames, Iowa: Iowa State College Press, 1946.
24. THURSTONE, L. L., & THURSTONE, THELMA G. *Examiner manual for the SRA primary mental abilities*. (Preliminary Ed.) Chicago: Science Research Associates, 1947.
25. THURSTONE, L. L., & THURSTONE, THELMA G. *Examiner manual for the SRA primary mental abilities (intermediate)*. (Rev. Ed.) Chicago: Science Research Associates, 1949.
26. VIGOTSKY, I. S. Thought in schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1934, **31**, 1063–1077.
27. WECHSLER, D. *The measurement of adult intelligence* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
28. WELLS, F. L. *Mental tests in clinical practice*. Yonkers, N. Y.: World Book, 1927.

Received Jan. 3, 1955.

EXPERIMENTAL EFFECTS OF EGO-DEFENSE PREFERENCE ON INTERPERSONAL RELATIONS[1]

ARTHUR R. COHEN

Yale University

A CENTRAL issue of contemporary social science has to do with the interrelationships between personality and social interaction. One important aspect of this problem is concerned with the quality of the interpersonal relations among different kinds of people. The present experiment is an attempt to relate the personality defenses of interacting individuals to a range of perceptions they may have concerning their interaction.

According to the most general dynamic interpretation, defenses are ways of handling unacceptable impulses which are striving for expression. The question may be by-passed as to whether they are successful or unsuccessful by viewing them as enabling the individual to function in the social world by turning his unacceptable impulses to channels of expression other than direct ones. Fenichel (6) distinguishes a number of ego defenses, among which the following were selected for study: projection, regression, reaction formation, avoidance (the repression-denial family), and intellectualization (a form of isolation).

Individuals may differ in their disturbances related to various stages of psychosexual development and in the intensity of these disturbances. They may also differ, and differ independently, in their characteristic mode of defense against these disturbing impulses. Thus, each of the various psychosexual disturbances may be associated in a variety of individuals with each of the various defense patterns. Given two individuals with a disturbance in the same psychosexual area, their preferred defense in this area may be the same or quite different. Both may handle intense oral strivings, for example, by avoidance, or one may use reaction formation while the other projects his oral strivings.

At least in theory these defense patterns should influence interaction. If, for instance, the two individuals with oral disturbance are interacting and their oral strivings are aroused, their patterns of defense should have direct implications for the way they behave toward each other. Two avoiders may both bury their disturbing oral strivings and reach an amicable relationship with relative ease. On the other hand, if one avoids expression by reaction formation and the other expresses it by intellectualizing (i.e., by bringing up the oral strivings in discussion and isolating the affect connected with them), these two should develop a relatively tense and mutually hostile relationship. The simplest deductions from such reasoning would state that if the interaction of two persons centers around a psychosexual disturbance which they share, then similarity of defense should generate a different sort of relationship from that generated by dissimilar defense patterns.

It is also possible that certain dissimilar defenses, if complementary, could lead to a good deal of solidarity between two persons, whereas certain similar ones might produce considerable conflict. In this latter connection, psychoanalytic theory suggests that projection is particularly applicable. Freud (7) defines projection as the attribution to the external world of impulses unacceptable to the ego. If two people interact and both tend to project the same unacceptable, disturbing impulses, a potentially explosive relationship can be anticipated if these impulses are aroused by the content of their interaction. Both may be expected to express the disturbance and project it, each attributing his unacceptable impulse to the other, fighting against this impulse in the other, and tossing it back and forth in this fashion as long as the interaction continues in this context. These considerations generate the following major hypothesis: when

[1] This experiment was carried out at the University of Michigan as part of a larger project concerned with testing psychoanalytic theory under the general direction of Dr. Gerald S. Blum. The project was financed by a research grant M-659 from the United States Public Health Service. The author wishes to express his indebtedness to Dr. Blum for his invaluable aid. Thanks are also due to Dr. Alvin Zander and to Edwin Thomas, Harry Burdick and Charles Uhl for the stimulation they provided at the outset of the study, and to Dr. Justin Weiss for his critical appraisal.

two people who share a common psychosexual disturbance, which they tend to defend against by means of projection, interact concerning content which arouses this disturbance, they will experience more negative affect in their relationship than will people who share a psychosexual disturbance, but have differing defenses associated with that disturbance.

It may be (at least with regard to projection) that a characterization of defenses in interaction as congenial or uncongenial would be more accurate. Defenses might then be viewed as congenial or uncongenial whether they are similar or dissimilar. It should be noted that the concept of congeniality or uncongeniality of defense as it is discussed here is different from that of compatibility or incompatibility of defenses within a given person's hierarchy of defenses. Here the interactive factor is stressed: the congeniality or uncongeniality of defenses of two individuals who are involved in a specified social relationship.

The hostile or negative interactions which are expected to ensue when defenses are uncongenial are of a variety of orders, all stemming from a conception of threat in interpersonal relationships (3). When the two persons threaten one another, they may: (*a*) be relatively unattracted to the relationship, to the other person, and to the task around which they interact; (*b*) perceive the other as hostile, unfriendly, as not giving them social support, and as self-centered; (*c*) develop a negative self picture and feel that their partner is attempting to exert influence upon them; and (*d*) be unmotivated in the situation and disposed to avoid such situations in the future. Of course, all persons with uncongenial defenses may not respond in all these possible ways, but in general, the uncongenials may be expected to do so in this manner more than the congenials.

METHOD

To test the hypotheses suggested above, a research design was developed that: (*a*) provided for the interaction of persons with common conflicts and given defenses associated with those conflicts, (*b*) experimentally aroused the conflict, (*c*) focused the content of the interaction around the conflict, and (*d*) measured the individual's perception of a variety of aspects of that interaction.

The procedure and rationale for determining the degree and areas of psychosexual disturbance for each *S* and the nature of the defenses associated with his disturbances have been fully reported elsewhere (2). Here it will be sufficient to note that the Blacky Pictures, developed by Blum (1), were used to assess psychosexual conflicts. The Blacky test is a modified projective device consisting of a series of eleven pictures which can be shown to individuals or groups. Each picture is geared to a major object relationship treated by psychoanalytic theory. From the *S*'s responses to the pictures, one may derive a series of scores indicating the degree of disturbance in such areas as oral, anal, oedipal, love object, superego, and so forth.

The major source of data for the assessment of defense was the Defense Preference Inquiry (DPI) for the Blacky Pictures. Some additional procedures were also used to assess defense but they were of relatively minor importance. The DPI taps defensive reactions to psychosexual stimuli in an indirect manner by having the *S* judge a series of alternatives in terms of the degree to which they represent the way "Blacky seems to be feeling or acting" in a particular picture (8). This encourages the *S* to identify with Blacky and thereby reveal his own personal reactions. Each alternative is an operational definition of a defense mechanism. The subject is asked to rank-order a given set of alternatives for each psychosexual dimension in a very short time, thus facilitating spontaneity of response.

For each *S*, then, a catalogue of his psychosexual disturbances and associated defenses was obtained. The next task was to group the *S*s in some arrangement that would allow the testing of the hypotheses as well as maximize the possibility for new and unexpected findings. It was decided to confine the interaction to two people; pairs are the simplest interpersonal model and in addition are relatively easy to manipulate experimentally. The design also called for limitation of interaction to people paired on the basis of the same psychosexual dimension, thereby controlling the variability which might result from differences in area of disturbance.

The main criterion for pairing was directly geared to the hypotheses of the study. The main hypothesis specified a difference between pairs of projectors and other pairs, and the secondary one a difference between similar and dissimilar defenders. Therefore three general types of pairs were decided upon: pairs of projectors, pairs of other similars of all kinds, and pairs of dissimilars of all kinds.

Since the experimental arrangement provided an opportunity for testing the effects of such other factors as dimension and intensity of psychosexual conflict, these became additional criteria for pairing. Five dimensions were chosen; the choices were determined by the representativeness of the dimension in psychosexual chronology as well as by its availability in the experimental population. For some dimensions, there were very few *S*s who had a given degree of conflict and an associated defense, and this precluded their use. The five psychosexual dimensions decided upon were: oral sadism, anal expulsiveness, oedipal intensity, castration anxiety, and sibling rivalry. Each *S* was used just once; he was paired with his partner on the basis of one of these dimensions, and one only.

TABLE 1

ARRANGEMENT OF EXPERIMENTAL PAIRS ACROSS PSYCHOSEXUAL DIMENSIONS, INTENSITY OF CONFLICT, AND TYPE OF DEFENSE PREFERENCE

Psychosexual Dimension	Similar Defenses (Both have +)	Similar Defenses (Both have 0)	Similar Defenses (One has +, the other 0)	Dissimilar Defenses (Both have +)
1. Oral sadism	+Av, +Av +RF, +RF +Rg, +Rg		+P, 0P +P, 0P	+Av, +Rg
2. Anal expulsiveness		0RF, 0RF 0RF, 0RF	+RF, 0RF	+RF, +Rg +Av, +I
3. Oedipal intensity	+Av, +Av			+P, +Rg +P, +RF
4. Castration anxiety	+Av, +Av		+P, 0P	+Av, +I
5. Sibling rivalry	+P, +P +P, +P		+Rg, 0Rg +RF, 0RF	+P, +Rg

Furthermore, each *S* had a certain degree of disturbance on the dimension on which he was paired. Some people had relatively little (indicated by 0) and others had a relatively intense disturbance (indicated by +). The projector, similar and dissimilar pairs were arranged so that in some cases both partners would have a high degree of disturbance (+, +) on the psychosexual dimension on which they were paired. In other cases, both partners had relatively little disturbance (0, 0). In still others one had a high degree of disturbance and the other relatively little (+, 0).

The final and limiting criterion for pairing was a sociometric one. No two *Ss* were paired if they had indicated one another as friends in a preliminary sociometric questionnaire. This procedure allowed some control over the effect of prior attitudes and past contacts.

The complete pairing arrangement for all *Ss* is given in Table 1.[1] Av designates avoidance; RF, reaction formation; Rg, regression; P, projection; and I, intellectualization.

Subjects

The *Ss* were 44 undergraduate members of a social fraternity at the University of Michigan who had made themselves available for a large program of testing and experimentation in connection with the psychoanalytic theory project. They were extremely homogeneous with regard to their interests and their educational, ethnic, and socioeconomic backgrounds.

Procedure

The *Ss* were run in three large groups of 16, 20, and 8. The experimental setting for all groups was a rather large lecture hall over which the *Ss* were told to spread out in pairs, one member of a pair taking a seat directly behind his partner.

The general experimental manipulations involved the following sequence:

1. ***The creation of a state of high motivation in the subjects.*** The *Ss* were told that they would be asked to interpret some brief descriptions of the behavior of some individuals. The importance of the task as an index of sensitivity and intelligence was impressed upon them and they were emphatically urged to try as hard as they could to do well. They were also told that the experimenter was very interested in their performance.

2. ***The arousal of the psychosexual disturbance in each partner separately so that the defenses associated with the disturbance would be stimulated.*** Each *S* was given a series of written stories concerned with the psychosexual dimension on which he was paired. For each dimension three stories calculated to tap the disturbance were provided. One located the particular disturbance in a home context, one in a school context, and one in the context of a social group. This was done because a representative selection of experiences seemed more likely to arouse the disturbance and stimulate the defense.

3. ***The heightening of the disturbance by the motive interpretation task.*** After the *S* had read the three conflict-arousal stories, they were collected and he was handed an "individual motive interpretation" blank. By having him write interpretations of the behavior of the "hero" of the stories when the behavior was directly concerned with the content of the *S*'s known psychosexual disturbance, his disturbance was expected to be heightened and the associated defense stimulated. The motive interpretation blank contained two questions: "Why do you think _____ acted this way in this situation?" and "What do you think are the underlying personality forces behind his behavior?"

4. ***Interaction of the partners: their involvement by discussion of the disturbance-arousing material.*** After finishing the individual motive interpretations the partners were instructed to discuss the material for 15 minutes, preparatory to writing their common interpretations. They were also apprised of the fact that observers would be watching closely to gauge the quality of their discussions. The observer device was instituted in order to keep motivation high throughout the discussions.

5. ***Interaction of the partners: their further involvement by having to work together over the disturbance-arousing***

[1] It should be noted that the possibilities for pairing were quite limited by the size of the sample on whom the personality assessments were obtained.

material. To further heighten the interactive aspect, the *S*s were asked to produce a set of joint interpretations. They were given motive interpretation blanks identical to the individual ones, except that it was indicated that this was to be a team effort.

6. *The measurement of their perception of their interaction*. The dependent measures were all assumed to tap manifestations of threat in interpersonal relations. In most cases they involved ratings by the *S* on a series of seven- and eight-point a priori scales. Others involved scaling devices such as "pick two," and one measure gathered the *S*s' responses into simple categories. Sixteen measures in all were used; they can be seen in Table 2.

RESULTS

Projection Hypothesis

The major hypothesis specified a difference between projectors and all other pairs, both similar and dissimilar. Projectors were expected to have more negative and hostile interpersonal relationships than other pairs. The results confirm this hypothesis, as shown in Table 2. In this table, the mean response on each dependent item for the *S*s who experienced the three main experimental arrangements is given; the lower the mean the more negative the interaction. The projector pairs are given in the middle column, with the dissimilars and similars on either side. The p values for the significance of the difference between the two other conditions and the projectors are given in the two columns between the three conditions. The t test was used on all measures but the last one, where the Exact test was employed.

The results show that the pairs of projectors are far more negative concerning their interaction than are those whose defenses are also similar but are other than projection. They are also more negative about their interaction than are those with dissimilar defenses. On almost every index of interpersonal relations, they perceive a more hostile atmosphere than either of the other two groups. They perceive their team's success on the task to have been minimal, feel that they were members of a poor quality team, and even think that other people do not do well on such tasks. They are less attracted to the interpersonal situation in general, and to the task and to their partner specifically. They perceive less support from their partner, perceive him to be more hostile and more interested in his ideas only, and are more insecure in the relationship. They feel that he is attempting to influence them more and also that they themselves tried to exert more influence. They withdrew their motivation from the task presumably as a self-protective device. And finally, they prefer to have

TABLE 2

PROJECTOR PAIRS VS. OTHER DEFENSE PAIRS (SIMILAR AND DISSIMILAR) FOR EACH DEPENDENT ITEM

Dependent Items	Mean of all Individuals in Similar Pairs (++, +0, 00) (N = 20)	p Value of Difference Between Similars and Projectors	Mean of All Individuals in Pairs of Projectors (++, +0) (N = 10)	p Value of Difference Between Dissimilars and Projectors	Mean of all Individuals in Dissimilar (++) Pairs (N = 14)
1. Attraction to interpersonal situation	9.1	<.001	5.6	<.001	9.3
2. Attraction to task	5.3	.08	4.2	.08	5.2
3. Attraction to other	10.3	<.001	6.4	<.001	9.4
4. Perceived success of team	4.6	<.001	3.0	<.001	4.9
5. Perception of individual success	4.1	ns	4.1	ns	4.4
6. Perception of others' success	3.4	.05	3.0	<.05	3.6
7. Perception of social support	4.2	.03	3.3	.05	3.8
8. Perception of partner's hostility	6.1	<.001	4.2	<.01	5.5
9. Perception of partner's self-interest	4.1	<.01	2.8	<.01	4.0
10. Security in relationship	5.6	<.001	3.5	<.001	5.9
11. Partner's attempted influence	5.3	.05	4.0	.04	5.3
12. Own attempted influence	5.0	<.01	3.2	<.01	5.5
13. Perceived quality of team	6.3	<.001	3.6	<.001	6.4
14. Desire to avoid such experiences	5.4	ns	5.2	ns	5.6
15. Motivation to do well	6.3	.04	5.2	<.05	6.2
16. Desire to organize work:					
independent	0		4		0
cooperative					
extremely cooperative	20	<.02	6	<.08	14
dependent					

the work organized in a way that emphasizes more independence. These results provide verification for the major hypothesis concerning the projectors.

Further Exploration of the Effects of Projection[3]

1. *Paired projectors vs. projectors paired with other defenders.* It appears that simply being a projector may not lead to so much negatively laden interaction. On all 16 dependent items, the projectors who were paired with other projectors experience their interaction as more negative than do the few projectors who were paired with others of a different defense. It may be expected then that if one is a projector, more hostile interpersonal relationships tend to ensue when the other person is also a projector.

2. *Effects of intensity of disturbance within projection pairs.* The group of paired projectors included pairs where both had relatively high disturbance (+ +) on the psychosexual dimension on which they were paired and pairs where one partner had high disturbance and the other low (+ 0). The data show that there is a trend for the pairs where both have high disturbance to perceive a more hostile interrelationship than the pairs where one partner is relatively disturbance-free. The +, + group reported more negative interaction on 10 out of the 16 items, the +, 0 group on four. They were equal in threat on two measures. It seems that even among the projector pairs, those with more disturbance tend to have more negative interaction.

3. *Effects of intensity of disturbance in projector pairs vs. pairs of other defenders.* This comparison was made to check on the possibility that it was only the projector pairs where both partners were highly disturbed which caused the entire projector group to be more negative. Accordingly, the pairs where only one member had a high degree of psychosexual disturbance were compared with all other groups where only one member was highly disturbed. The trends suggest that the paired projectors where only one partner has a high degree of psychosexual disturbance have more hostile interaction (more threat on 11 items) than do other pairs where only one partner is relatively disturbed (more threat on four items; one tie). Thus, though the former are less threat-oriented than the pairs where both projector partners are relatively disturbed, they still are more so than the rest of the population.

4. *Projectors paired with other defenders vs. dissimilar defender pairs excluding projectors.* Finally, an exploration was undertaken of the difference between the projectors who were paired with someone else and other members of dissimilar pairs to see if projection alone led to negative interaction. The data indicate that the projectors when not paired with other projectors are not more threat-oriented than other dissimilar pairs.

It was seen earlier that paired projectors have more negatively toned interaction than other similar and dissimilar pairs of all kinds. The data which followed suggest further that it may not be projection alone, but being paired with another projector which makes the difference. The pairs tend to be more negative than those who are alone, and the latter are not more negative than other people who do not have partners with identical defenses. Furthermore, there appear to be differences even among the paired projectors: when both partners have a high degree of psychosexual disturbance, more threat tends to be produced than when only one is highly disturbed. Finally, paired projectors with one having high disturbance are more negative in their perceptions than are pairs of other defenses where only one partner is disturbed.

Similarity-Dissimilarity Hypothesis

The possibility was also raised that people with dissimilar defenses associated with a common psychosexual disturbance might perceive their interaction differently from those with similar defenses. Accordingly, all those

[3] These analyses and the succeeding explorations were carried out by taking into account direction of difference, not magnitude. The comparisons between two given conditions were made by taking their means on all the dependent items and seeing which group had lower means (an indication of more hostile interaction). By pointing to trends in this manner, it was possible to isolate some of the more specific conditions of defense and disturbance under which negative interpersonal relationships take place. The important point here is the consistency of the trends; the preponderance of means indicating more negative interaction among certain groups suggests that the given group may be behaving differently from another group and that this difference is meaningful. p values for sign tests could be given here but it was felt that the question of the independence of measures of interaction could be raised. Therefore, the more conservative course of inspecting the distribution of means was taken.

TABLE 3

COMPARISON OF VARIOUS DEFENSE PAIRS WITH EACH OTHER

(Projection, $N = 10$; Regression, $N = 4$; Reaction Formation, $N = 10$; Avoidance, $N = 6$)

Projection vs. All Others	No. of Dependent Items	Regression vs. Others	No. of Dependent Items	Reaction Formation vs. Avoidance	No. of Dependent Items
P's show more threat	16	Rg's show more threat	14	RF's show more threat	14
Av's show more threat	0	Av's show more threat	2	Av's show more threat	2
P's show more threat	16	Rg's show more threat	11		
RF's show more threat	0	Rg's and RF's equal	1		
		RF's show more threat	4		
P's show more threat	13				
P's and Rg's equal	1				
Rg's show more threat	2				

pairs with similar defenses (except for projectors) were compared with those with dissimilar defenses on each of the 16 dependent measures. The means for the similar group are given in Table 2, column 1, for the dissimilars in Table 2, column 5. When these comparisons were made, no differences whatsoever were found; both groups perceived their interaction to have been relatively smooth and rewarding.

Further Explorations of Disturbance and Defense

1. *Comparison among Blacky dimensions in terms of perceived negative interaction.* Of further interest were the differences between people who were paired on the basis of different psychosexual dimensions. Disregarding specific defense and intensity of disturbance, one may also explore the effects of the various dimensions themselves. Comparisons on the 16 dependent items were made between those people who received oral sadism stories and those who received anal expulsiveness stories, oral sadism stories versus oedipal stories, and so on.[4] Ten comparisons were made; every dimension with every other, using the sign test model. They are too lengthy to be reported here in full and may be summarized.

Briefly, the dimensions can be ordered according to the degree to which people paired on them experienced negative interpersonal relationships when interacting around relevant disturbing material. It appears that those paired on sibling rivalry had the most negatively toned interaction, compared with those people who were paired on the other four dimensions. On the other hand, pairing on oedipal intensity was related to the least threat-oriented interaction. The order determined by these 10 comparisons was as follows, from most negative to most positive:

1. Sibling rivalry
2. Castration anxiety
3. Oral sadism
4. Anal expulsiveness
5. Oedipal intensity

Of course, the dimensions in the middle range shade into one another, but at the extremes the dimensions are clearly different in their evocation of interpersonal hostility. It may also be that the projector pairs are responsible for the negative effect of the sibling rivalry dimension since two pairs fall in this area. However, it is quite possible that sibling rivalry occupies the place it does in the hierarchy independently of projection. This is an interpersonal situation where two undergraduate peers are interacting around disturbing sibling rivalry content.[5]

[4] The N's on these dimensions are: sibling rivalry, 10; anal expulsiveness, 10; oral sadism, 12; oedipal intensity, 6; castration anxiety, 6.

[5] The question may also be raised as to the general effect of sibling rivalry. Is it possible that just having sibling rivalry may lead the individual to experience his relations with a peer as negative? If this is true, then perhaps it is disturbance on the dimension of sibling rivalry and not a projector paired with another projector which is generating differences in the perception of interaction. To answer this question, a differentiated analysis of sibling rivalry was undertaken. These results show that the mere presence of sibling rivalry does not seem to account for the differences found previously. It appears that those who had a disturbance on the dimension and also interacted around disturbance-arousing sibling rivalry material were more negative concerning their interaction than those who did not receive the stimulus, whether they had the disturbance

2. *Hierarchy of defense preference according to perceived negative interaction.* Table 3 summarizes the comparisons made between the different types of defense pairs. In this analysis only pairs of similars were used; there were too few different types of dissimilars to permit their exploration. Intellectualization was omitted because there were no similar pairs using this defense.

The data strongly suggest a hierarchy of defenses in terms of their effect on the perception of interpersonal relationships. Pairs of projectors appear to be more negative concerning their interaction than pairs who had other defenses associated with psychosexual disturbances. This reflects the confirmation of the main hypothesis. Furthermore, regression appears to bring more negative interaction than reaction formation and avoidance, and reaction formation itself more than avoidance. The order is as follows from most negative to most positive:

1. Projection
2. Regression
3. Reaction formation
4. Avoidance

Thus, it appears that among people whose interaction is colored by a common psychosexual disturbance and who have similar defenses associated with that disturbance, there are distinct differences depending upon the defense. In general the defenses which stress the avoidance of an unacceptable impulse (avoidance and reaction formation) generate more cohesive interaction than those which permit the impulse to gain some sort of expression (projection and regression).

DISCUSSION

The results of this investigation raise a number of interesting problems. It was seen that the hypothesis regarding the interaction of projectors was confirmed. The reasoning behind this hypothesis assumed that projectors in interaction around disturbing material would keep it in the open as a source of constant anxiety, thereby generating hostile interpersonal relationships. The person who uses projection (as well as any other defense) in conjunction with a given psychosexual disturbance is fighting against an unacceptable impulse. In the present situation he is in interaction with a partner who gives him a perfect opportunity for fighting the impulse by having it himself. And since the projector's characteristic mode of warding off unacceptable impulses is to perceive them in another person, the other person may become the symbol of these impulses. In effect, the projector may be said to fight against the impulse by fighting the other member of the pair.

Fenichel notes that the paranoid person who uses projection "is sensitized, as it were, to perceive the unconscious of others, wherever this perception can be utilized to rationalize his own tendency toward projection" (6, p. 147). The present experimental situation made it relatively easy for an individual to see his own unacceptable impulses in the other person, and to react negatively to them. This was expected, and found, to be easiest for the projector pairs, where the same processes obtained for both partners. In this manner, a mutually hostile interpersonal relationship was generated.

This line of reasoning also permits an explanation of some of the additional findings concerning projection. It was seen that more negative interaction resulted when both partners were projectors than when only one member of a pair used this defense. Here, of course, the fact that both partners are behaving in the fashion described above tends to increase the interpersonal conflict; when only one partner behaves this way, he may not receive so much additional support for his behavior through his partner's actions.

One may also view in this light the finding that more negative interaction ensues when both projector partners have a high degree of psychosexual disturbance than when only one member of the pair is highly disturbed. What may be happening here is that the defense of projection is not operating too effectively for the highly disturbed people and they may be driven to this interpersonal conflict as a way of working out the defense. It may also be that when both partners have strong conflict the interaction will produce more expressions

or not. On the other hand, those who had the disturbance and did not get the sibling rivalry stories were no different from those who were free from sibling rivalry problems and did not get these stories. It may be said then that disturbance on this dimension is meaningful only when there has been some arousal of the disturbance.

of the unacceptable impulse, thereby increasing the threat and the hostility and facilitating further projection. When both partners are engaged in such a struggle, the conflict may be raised to a higher pitch than when only one acts in this manner. It should be emphasized that the foregoing is largely speculative; it is, however, one way of explaining the present results.

It was also seen that while the projection hypothesis was confirmed, the secondary hypothesis concerning differences between similars and dissimilars was not. Though limited to the field of projection, the present evidence points to a theoretical position with regard to the interaction of people with given defenses which favors the congeniality-uncongeniality notion rather than that of similarity-dissimilarity. While groups of people with dissimilar and similar defenses other than projection reacted equally favorably to their experiences in this experimental situation, those groups of similars where both were projectors perceived their interaction in a more negative and threat-oriented fashion. Apparently, the use of such an expressive defense as projection by two individuals in interaction causes the maintenance and heightening of the disturbance and leads them to threaten one another. Other sorts of expressive defenses, as well as avoidance defenses which block the outlet of unacceptable impulses, whether people have them in common or in different combinations, appear to enable them to handle the disturbing material in a more efficient and rewarding manner.

The results also indicate the presence of a hierarchy of defenses with regard to perceived negative interaction among partners. The defenses associated with the most negative interaction were seen to be projection and regression in that order. These are defenses permitting the unacceptable impulse to gain an outlet. In the present situation such a mode of defense leads to many interpersonal complications. On the other hand, the avoidance mechanisms which appear to lead to smoother interaction are more easily accommodated by the experimental task. Those partners who use these defenses can avoid unpleasantness by virtue of their tendency to repress, deny or ignore disturbing material. Each, thereby, contributes to the increasing security of the other in the relationship.

Further evidence consistent with these findings is available. In a related investigation of personality and sociometric choice, within the same research project, the author (4) found that people who used projection against given disturbances underchose others who also used projection against these same disturbances. No other significant underchoices were found. In the same study, those people who used avoidance defenses were seen by their social group to be less deviant from the norms and values of that group. In addition, the use of the avoidance defenses was found (5) to be related to high self-esteem, whereas projection tended to be associated with low self-esteem. Thus, in an open friendship choice situation, and on a self-rating questionnaire, the different defenses appear to have differential consequences for social adjustment.

SUMMARY

This experiment was designed to explore the connection between the personality defenses of interacting individuals and their attitudes toward and perceptions of their interaction. The reasoning behind the experiment assumed that two people in interaction who had a psychosexual disturbance in common would react to one another, when that disturbance was aroused, as a function of their defenses against that disturbance.

The assessment of psychosexual conflicts and ego-defense preferences specific to each individual was made prior to the experiment. The Blacky Pictures technique and its auxiliary Defense Preference Inquiry were used for this purpose. Five psychosexual dimensions were used: oral sadism, anal expulsiveness, oedipal intensity, castration anxiety, and sibling rivalry. The following defenses were studied: projection, avoidance, regression, reaction formation, and intellectualization.

The *Ss* were paired in terms of defense, psychosexual dimension and intensity of disturbance. The limiting criterion for pairing was a sociometric one which assured some control over the affective ties among the *Ss*. The basic pairs were of three kinds: pairs of projectors, pairs with similar defenses other than projection, and pairs where the defenses were dissimilar.

It was found that:

1. When in interaction involving a task which arouses a specific psychosexual disturbance they have in common, two people who tend to project this impulse will experience their interaction as more negative than pairs of people who utilize other defenses.

2. The mere presence of projection however, may not be sufficient to lead to interpersonal hostility. Negative effects seem to be obtained only when two projectors are paired. Projectors interacting with people using other defenses tend to be no different from any other pairs having dissimilar defenses.

3. There may be differences among the pairs of projectors depending upon whether one partner or both has a high degree of psychosexual disturbance on the dimension along which they were paired. When both partners have high conflict, more negative interaction tends to result than when only one is highly disturbed.

4. Similarity or dissimilarity of defense per se appear to make no difference as far as quality of interpersonal relations is concerned.

5. Among those with similar defenses however, there appears to be a hierarchy of efficiency with regard to interpersonal relations. Pairs of avoiders reported more positive interaction than pairs of other defenses. Reaction formation was next best, regression next, and projection worst in this regard.

6. The psychosexual dimensions themselves were ordered in terms of the degree to which they stimulated negative interaction. The order from most negative to most positive was: sibling rivalry, castration anxiety, oral sadism, anal expulsiveness, and oedipal intensity.

REFERENCES

1. BLUM, G. S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, **39**, 3–99.
2. BLUM, G. S. Procedure for the assessment of conflict and defense. Unpublished manuscript, Univer. of Michigan, 1954. (mimeographed)
3. COHEN, A. R. Situational structure and individual self-esteem as determinants of threat-oriented reactions to power. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
4. COHEN, A. R. Personality and sociometric choice. Unpublished manuscript, Univer. of Michigan, 1954. (mimeographed)
5. COHEN, A. R. Some explorations of self-esteem. Unpublished manuscript, Univer. of Michigan, 1954. (mimeographed)
6. FENICHEL, O. *Psychoanalytic theory of neurosis.* New York: Norton, 1945.
7. FREUD, S. Psychological notes upon an autobiographical account of a case of paranoia. In *Collected papers*, London: Hogarth, 1925.
8. GOLDSTEIN, S. A projective study of psychoanalytic mechanisms. Unpublished doctoral dissertation, Univer. of Michigan, 1952.

Received January 5, 1955.

THE EINSTELLUNG TEST AND PERFORMANCE IN FACTUAL INTERVIEWING

KURT W. BACK[1]
University of Puerto Rico

THE interviewer who must collect a wide range of information from a variety of informants is subject to numerous pressures. In order to carry out his job effectively, he must follow detailed instructions on what information he is to obtain, and he must be able to adjust his style to suit each particular interview. Consequently, the conditions under which he maintains and under which he changes his routine can be important factors in his performance. Hence, a test in which the subject (*S*), while complying with instructions, is given the opportunity to adapt a previously established pattern to new conditions is likely to measure relevant personality variables.

The water-jar problems were used as probably the best available predicter for the purpose. It has been used previously by Luchins, Rokeach, and others as a predictive measure (3, 5, 8), but its use as a test of rigidity and the whole concept of rigidity as a personality trait has recently been shown to be of doubtful value (1, 4). It was felt, therefore, that it was necessary to make a new start in interpreting performance on this test. In this study, the water-jar problems were used, not to measure rigidity, but merely to show the various conditions under which a person will maintain a set (or apply a learned solution). Hence, it could be compared to the conditions under which, as an interviewer, he would act in an analogous manner.

PROBLEM

The study was conducted in a large-scale information-collection program for the U. S. Air Force, in which the interviewers conducted lengthy, detailed interviews about the characteristics of foreign areas. The outlines of the interviewers' reports and the topics to be covered were prescribed in great detail. The manner of interview, the exact questions to be asked, and the general approach to the respondent were left to the interviewer. Under these conditions, personality tendencies toward maintaining a set or adapting to circumstances can easily affect the interviewer's style of work and thus his effectiveness. The aim of this study was to determine the relationship of the water-jar test as a measure of this trait to the interviewer's performance.

METHOD

Subjects

The *S*s of the study were interviewers working in Germany and Japan. Their jobs consisted of interviewing German, East European, and Japanese nationals, respectively, who had relevant information on foreign areas, and of writing standardized reports on the data they had collected. The German group included German nationals and about 10 per cent U. S. civilians (primarily of German origin). The Japanese group was made up of U. S. citizens of Japanese origin. In both groups the age ranged from 20–60 and the educational level from high school to doctorate, most of the interviewers having had some college.

The number of *S*s available for different parts of the study varies somewhat, as not all measures could be obtained for all *S*s because of turnover. The number ranges from 156–205 in Germany, and from 39–53 in Japan.

No claim can be made that either group is representative of its culture. However, the replication gives it some cross-cultural generality.

The Test

The form of the test was adapted from the one designed by Luchins (5). Its outline is shown in Table 1. The five "training problems" are intended to establish a set toward the solution B-2A-C. Only *S*s who established the set were considered in the study. The criterion for establishing was solving at least either three of the five problems or the last two problems by the standard method. In Problems 6 and 7 (critical A), the same solution will solve the problem but a shorter solution is possible. Problem 8, the "extinction problem," cannot be solved by the original method, but only by the short method. The last two problems (critical B) can be solved by both methods.

[1] This paper represents part of a study conducted at the Bureau of Applied Social Research, Columbia University, under Contract 33(038)-14313 with Human Resources Research Institute, Air Research and Development Command, Maxwell Field, Alabama. The contributions of all members of the staff of this study are gratefully acknowledged. In particular, Lee Wiggins, Robert Beezer, Belle Buchman Wiggins, and Bernard Levenson had a major part in the collection and analysis of the data used in this paper. This may be identified as publication No. A-183 of the Bureau of Applied Social Research, Columbia University.

TABLE 1

THE WATER-JAR PROBLEMS AFTER LUCHINS (3)

Problem No.	Type of Problem	Jar			Total Required
		A	B	C	
1	Training	21	127	3	100
2	Training	14	163	25	99
3	Training	18	43	10	5
4	Training	9	42	6	21
5	Training	20	59	4	31
6	Critical A	23	49	3	20
7	Critical A	15	39	3	18
8	Extinction	28	76	3	25
9	Critical B	18	48	4	22
10	Critical B	14	36	8	6

In most studies *Ss* are classified as rigid if they persist in the long solution after it is not necessary any more (Problems 6 and 7) and especially after it was shown to be inapplicable in one case (Problems 9 and 10). As Luchins (6) has pointed out, it may be easier to keep using a proved, easily applicable method than to keep looking for new conditions. In the present case we are less interested in measuring a hypothetical personality variable called "rigidity" and even less in deciding its value in problem solving, than in relating the performance on this test to the *S*'s ability as information-collecting interviewer. In approaching this question, we can classify a generalized attitude of a person who has found a certain approach effective in several situations and is confronted with similar situations. The following are possible ways of acting:

1. He can accept the learned solution tentatively, ready to change when difficulties arise. He is striving for perfection, making the effort to pay attention to the requirements of each situation. We shall call him the *situational type.*

2. (*a*) On the contrary, he can trust the learned solution to guide him safely in the future, accepting the risk of acting inappropriately in a particular situation as the price for ease and speed of working. This type stays quite independent of the particular situation (*nonsituational*), and restricts his repertoire to few responses. We shall call him the *small-repertoire type.*

2. (*b*) He can stay independent of the particular situation by accepting alternative ways of proceeding as part of the requirements of the total situation. He may change his ways of action, not because of the requirement of the instant, but on the basis of his own decision. He is also *nonsituational*, and we shall call him the *large-repertoire type.*

The performance on the *Einstellung* test can be used as an indication of the approach. The situational type can be expected to shift to the short solution when it is necessary, i.e., at the extinction problem. The small-repertoire type will not attempt to introduce the short solution at all, accepting or not noticing the failure at the extinction problem. The large-repertoire type will correspond to the *S* who introduces the short solution, but not from necessity (i.e., not at the extinction problem) or even returns to the long solution after using the short one.

TABLE 2

THE PRINCIPAL PATTERNS OF SOLUTION

Problem	Situational	Nonsituational	
		Small Repertoire	Large Repertoire
6	Long	Long	Long or Short
7	Long	Long	Long Short
8	Short	Long or no solution	Short Short
9	Short	Long	Long Short
10	Short	Long	Long Short

Table 2 shows the major patterns of solution in each type.

Criteria of Interviewer Performance

In the course of the study, several measures were used as criteria of interviewer performance. The four of them used here measure principally performance in collecting the required information (as distinguished, for instance, from writing up the reports and other clerical skills). Two of these criteria were based on supervisors' ratings, combined on the basis of a factor analysis, and two on content analysis of a number of the interviewers' reports. These were:

Compliance with instructions. This index was based on supervisors' ratings of traits associated with obtaining the required information, such as use of interviewing guide, inclusion of adequate details, and thorough exploitation of informant's knowledge.

Ability to deal with people. This index, also based on supervisors' ratings, includes several ratings of interpersonal effectiveness, such as skill in making the informant feel at ease, over-all interviewing ability and supervisory potential.

Quantity of information. This index was based on content analysis of the interviewers' reports and consisted not merely of the gross quantity of information in the reports, but, in addition, of such measures as number of different topics and number of different types of content.

Efficiency of obtaining information. This index was based on the ratio of the amount of information in the reports of an interviewer to the time consumed by the interviewer, both in interviewing and report writing.

Intercorrelation of these indices ranged from .20 to .76, with a median of .29 in Germany; and .16 to .76, with a median of .39 in Japan.[2] They measure related, but by no means identical, aspects of the interviewer's work.

Self-Ratings of Performance

During the course of the study, the interviewers completed a check list of evaluations for a number of their

[2] The complete matrices of intercorrelation are available in the technical reports on the study (tentative title: Wiggins, L. and Back, K.: *Studies in Factual Interviewing*) to be released by Air Force Personnel and Training Research Center, Air Research and Development Command, Maxwell Air Force Base, Ala.

interviews. Among other questions they rated their use of 18 techniques (such as "Talked to the informant about personal problems" or "Acted distant but polite") on choices of "never use," "use sometimes but not this time," "used less than usual," "used the same as usual," "used more than usual." The proportion of answers of "not this time," "less than usual," and "more than usual" was taken as a measure of the interviewer's variability from interview to interview.

RESULTS

The criteria measure principally the efficient obtaining of required information. This corresponds to the realization that in large-scale interviewing organizations, a premium is set on planning and control by the interviewer and little stress is put on subtle techniques, deep interviewing, and the respondent's reaction.

Under these conditions, the interviewer who adjusts to each short-range situation within the interview (e.g., changes in the informant's attitude) will generally be inferior to the one who sticks to a design. In addition, the interviewer who follows an over-all plan, but varies between several methods, will be superior to the one who relies on only one system. Applying this reasoning to the categories developed in the Einstellung test, we have the following rank order: large repertoire, small repertoire, situational.

Table 3 shows that the results tend to confirm the reasoning. The rank order on all four criteria is as expected, in spite of the low intercorrelations between the criterion measures themselves. Furthermore, the rank order is identical in both Germany and Japan. Thus, in eight different (although not independent) evaluations of performance we find the large-repertoire interviewer superior, followed by the small-repertoire and the situational interviewer.

In Germany some of the differences between groups reach an acceptable level of statistical significance. That more do not is to be expected in a study done in an operating organization. Randomization of respondent assignments was not practicable, and also organizational conditions affected efficiency. These were only a few of the extraneous factors interposed between the work of the interviewer and the evaluation of his performance. The consistent picture of the direction and size of the differences in interviewer criteria gives confidence in the interpretation of the Einstellung test and its application to the effectiveness of interviewers.

In the interpretation of the Einstellung test as a measure of rigidity, both the situational and the large repertoire groups would be more "flexible" than the small repertoire group. The situational interviewer, however, is influenced strongly by external conditions, adapting to the respondent and hence his methods will change from interview to interview. The large repertoire interviewer, however, will try most of his techniques in *each* interview and hence will vary little between interviews. Table 4 shows how the different interviewers describe the variability of their behavior.

TABLE 3

EFFECTIVENESS OF VARIOUS TEST GROUPS

Criterion of Performance	Situational		Nonsituational						Significance of Differences	
			Total		Small Repertoire		Large Repertoire		Situational-Nonsituational	Small-Large
	Mean	*N*	Mean	*N*	Mean	*N*	Mean	*N*		
Germany										
Deal with people	3.13	52	4.00	150	3.75	102	4.52	48	$CR = 2.33$; $p < .05$	$CR = 2.03$; $p < .05$
Compliance	3.87	53	4.95	152	4.93	101	5.04	51	$CR = 2.36$; $p < .05$	$CR = 1.10$; n.s.
Quantity	4.45	40	4.78	120	4.72	80	4.88	40	$CR = 1.03$; n.s.	$CR = 1.16$; n.s.
Efficiency	2.39	53	2.76	120	2.67	80	2.92	40	$CR = 2.27$; $p < .05$	$CR = 1.32$; $.10 < p < .20$
Japan										
Deal with people	3.36	14	4.13	38	4.00	29	4.56	9	Differences in Japan not significant by themselves.	
Compliance	3.93	14	4.87	39	4.69	29	5.40	10		
Quantity	4.67	9	4.95	23	4.74	19	5.75	4		
Efficiency	1.67	12	1.76	29	1.75	25	1.80	5		

TABLE 4
INDEX OF SELF-RATING OF FLEXIBILITY OF DIFFERENT TEST GROUPS

Countries	Situational		Nonsituational					
			Total		Small Repertoire		Large Repertoire	
	Mean	*N*	Mean	*N*	Mean	*N*	Mean	*N*
Germany*	5.19	36	4.66	121	4.78	80	4.44	41
Japan	5.67	12	5.86	29	5.96	24	5.40	5

* Significance of differences: Situational-nonsituational CR = 1.66, $.05 < p < .10$; Small-large CR = 1.28, $p < .20$

Note.—A larger score indicates self-rating for greater flexibility.

In Germany the difference between the situational and large-repertoire groups is in the expected direction and is significant ($p < .05$), the small-repertoire group falling between. In Japan the results are not clear. This is partly accounted for by a higher general level of variability in Japan. The influence toward variation, be it cultural or organizational, may obscure the influence of personality traits.

The correlations between the self-ratings of variability and the criteria cluster around zero. It is the test which measures the basic traits needed for success, while the performance ratings attenuate this relationship.

DISCUSSION

The relationship between a person's performance in the Einstellung test and his effectiveness as an information-collecting interviewer lends support to our hypothesis that the test measures a general way of approaching a series of situations, and that this way is important for performance as an interviewer. It may be noted that the use of some measure of rigidity (e.g., the long solution in the last two problems) would have led to slight evidence in favor of the rigid interviewer: the situational interviewer would be flexible, the small-repertoire interviewer would be rigid, and some large-repertoire interviewers would be flexible and some rigid. By separating the so-called "flexible" and "rigid" *S*s on the basis of their approach to the whole test, we find traits that can be generalized to behavior in other situations.

The basic difference between the situational and the nonsituational solutions in the test corresponds to the difference between adaptation to each particular situation within the interview and emphasis on the ultimate outcome. The group which emphasizes the whole performance against each operation can be divided again by the manner in which success in the whole performance (test or interviews) is sought: by relying on a method which had been shown to have a reasonable chance of success or by understanding the purpose of the whole proceeding and being able to try out several methods in pursuing it.

The conditions of the interview favored the nonsituational interviewer. There was no fixed schedule to be followed, and, as each informant was to be interviewed only as a source of information, the length and scope of the interview was dependent on the amount of information. Both conditions favor the interviewer who keeps in mind mainly over-all production and does not worry about the form of each particular question or even the result of each particular interview. Moreover, the interviewer who can vary his methods at will should be more effective even in his approach to over-all good performance than the one who restricts himself to repeating his previous successes. It should be kept in mind that the criteria used depend mainly on the value and amount of the total information transmitted and not on evaluation of the details of the interviewing as such, and thus conform to the attitude of the nonsituational interviewer.

The limitation of the data on the type of criteria, based on long-run performance, is important and the findings cannot be generalized to all types of interviewing. On the contrary, the interviewer who is superior in factual, informational interviewing, where the success of the interview depends on the total information obtained, may be inferior in situations where results do not cumulate, e.g., in clinical or survey depth interviewing, or where a complicated pattern on the questionnaire has to be followed. For instance, it has been shown that the abilities connected with following a schedule faithfully tend to be correlated negatively with those corresponding to understanding of the purpose of the whole study (2).

Finally, it is worth noting that the three types distinguished here show a rough analogy to Riesman's (7) personality types. The situational type would then correspond to the "adjusted" who adapts himself completely to

the environment. The small-repertoire type corresponds to the "anomic" who clings to a few learned patterns, and the large-repertoire type corresponds to the "autonomous" who "are free to choose whether to conform or not." Whether a direct connection can be established between the Einstellung test and these types, must be left to further research.

SUMMARY AND CONCLUSIONS

1. The Einstellung (three-pitcher) test was used as a predictive measure of interviewers' performance in large scale, factual interviewing. The test was not assumed to measure a personality trait of rigidity, but the patterns of answers were classified in the following types according to the whole test pattern: situational and nonsituational, the latter divided into large repertoire and small repertoire.

2. On four criteria of interviewer performance, the nonsituational was superior to the situational interviewer, and the large-repertoire to the small-repertoire interviewer. This was true of a group of interviewers in Germany and a group in Japan.

3. The nonsituational interviewers varied less in their use of techniques from interview to interview, as reported by them, the large-repertoire interviewers varying least.

4. It was suggested that the test be interpreted as a measure of a general approach to new problems in the face of learned solutions instead of structural rigidity. It would not be surprising, therefore, if it does not relate to other measures of structural rigidity, as has been reported frequently, while the present treatment seems promising in predicting behavior.

5. In interpreting the performance of the interviewers, it was necessary to distinguish the type of work (informational, factual interviewing) where the respondent is needed only as a transmitter of information, from other types of interviewing (clinical, attitudinal) where the success of each interview has to be measured independently.

REFERENCES

1. APPLEZWEIG, DEE G. Some determinants of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1954, **49**, 224–228.
2. BACK, K. W., HILL, R., & STYCOS, J. M. Interviewer effect on scale reproducibility. *Amer. sociol. Review*, 1955, **20**, 443–446.
3. COWEN, E. C., & THOMPSON, G. E. Problem solving rigidity and personality structure. *J. abnorm. soc. Psychol.*, 1951, **46**, 165–176.
4. GOODSTEIN, L. C. Intellectual rigidity and social attitudes. *J. abnorm. soc. Psychol.*, 1953, **48**, 345–353.
5. LUCHINS, A. S. Mechanization in problem solving: The effect of *Einstellung*. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
6. LUCHINS, A. S. Rigidity and ethnocentrism: A critique. *J. Pers.*, 1949, **17**, 449–466.
7. RIESMAN, D. (in collaboration with Denny, R. and Glazer, N.). *The lonely crowd*. New Haven: Yale Univer. Press, 1950.
8. ROKEACH, M. Generalized mental rigidity as a factor in problem solving. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.

Received January 6, 1955.

ASSIMILATIVE PROJECTION AND ACCURACY OF PREDICTION IN INTERPERSONAL PERCEPTIONS[1]

RICHARD M. LUNDY[2]

The Ohio State University

THE broad area of the relationships between people has interested many psychologists. A more restricted area, that of considering this relationship in terms of interpersonal perceptions, has more recently become a topic of concern to social and clinical psychological research.

Within this area of interpersonal perceptions, those perceptions which are accurate and those which are the same as one holds toward himself define the problem area this research sets out to explore. We attempt to demonstrate that predictable results regarding accuracy and projection will follow the introduction of the variable of a person's focus of attention while he is interacting with another person.

THEORETICAL CONSIDERATIONS

The psychology of personal constructs. Kelly has formulated a system of thought about psychological data which he calls the psychology of personal constructs (5). Kelly attempts to understand a person's processes by understanding the ways in which that person anticipates events. Kelly calls these ways of anticipating, personal constructs.

The present research stems directly from Kelly's position and is concerned with ways in which the perception of other people is related to an individual's personal constructs and with variables affecting personal constructs.

Construing objects. One of the dimensions of conceptualizing, or construing, is incorporation-differentiation. An object being construed may be incorporated into the constructs about another object or objects in the surrounding field, or it may be differentiated from these objects. In the interpersonal situation in which two people are talking to one another, this dimension takes the form mainly of incorporating the other person into constructs one holds about oneself, or differentiating the other person from constructs one holds toward oneself.

We think of incorporation as the bringing of an element into a construct which has its principal focus of convenience in other elements. In the case of interpersonal perceptions, the constructs into which the other person, as an element, is incorporated are those constructs which have the self as a principal element.

Differentiation takes place when the construing of an object is such that the construct has its principal focus of convenience in events associated with that object. In the interpersonal situation, then, the differentiation constructs would have their principal foci of convenience in the events associated with the other person.

The focus of attention of the construing person influences which process will take place. If attention is upon the construer, then incorporation will be facilitated. The facilitation comes from the fact that the person is drawing constructs about himself into the context of his conversation with the other person. Differentiation of the other person from himself is reduced because the construing of the other person takes place at the time that the focus of attention highlights constructs about the self.

Focus of attention upon the other person facilitates differentiation because it highlights the events associated with the other person and thus aids the formation of constructs that have those events as their elements. At the same time, attention upon the other person tends to exclude constructs which do not have the other person as their focus of convenience. The application of constructs about oneself to the other person is thereby reduced.

The above statements imply the necessity of construing the other person while talking with him. We assume that to talk to a person at all one must place some construct upon him and the situation. If the other person is relatively unknown, and one is in a situation which calls for interpersonal communications,

[1] This paper is based upon a portion of a doctoral dissertation completed in 1954 at The Ohio State University under the direction of Dr. George A. Kelly.

[2] Now at the University of Wisconsin.

then the construction of the other person must continually undergo revision as one attempts to keep up with the level of interaction and to incorporate into some meaningful structure the various events which make up the situation.

The constructural change which, in a sense, is forced upon a person in the interpersonal situation is in the direction of incorporation of the other person into the constructs the construer holds about himself when he is paying attention to himself and is in the direction of differentiation of the other person from constructs about the self when the construer is paying attention to the other person.

HYPOTHESES

The hypotheses are all framed in terms of an interpersonal situation utilized by Bieri (1) in his study of assimilative projection. Two people who know one another only minimally—some of the *S*s were in the same class and had seen one another previously—are introduced to one another and asked to discuss a common topic for a few minutes. The hypotheses are made in regard to the interpersonal perceptual change which occurs between tests taken before and after the brief conversation.

Hypothesis I. In a situation in which *S* is focusing his attention upon himself he will increase the degree to which he projects his own constructs upon another person.

Hypothesis II is a control statement for the first hypothesis and has to do with the comparison of the two situations: the directional increase in the degree to which *S* projects his own constructs upon another person will be greater in a situation in which he focuses his attention upon himself than in a situation in which he focuses his attention upon the other person.

Hypothesis III. In a situation in which *S* is focusing his attention upon the other person, he will increase the degree to which he accurately predicts the other person. Because of the probability that in a peer interaction-situation *S*s increasing in projection might also increase in accuracy, this hypothesis was made with the condition that accuracy be measured with the effects of projection controlled.

Hypothesis IV. The directional increase in the degree to which *S* accurately predicts another person will be greater in a situation in which he focuses his attention upon the other person than in a situation in which he focuses his attention upon himself.

METHOD

Subjects. The *S*s were 52 students, 34 male and 18 female, in the beginning psychology classes at The Ohio State University. They were drawn from four different classes.

Test materials. The usual approach to measuring interpersonal perceptions by having the *S*s fill out a rating scale or questionnaire as they think another person filled his out was used in this research. In selection of the test, our concern was in the degree to which a certain set of scores matches another set. The meaning of the scores was not of primary concern. What we do assume is that if an *S*'s score on one of the test blanks is the same as his score on another, then he was using similar constructs in his decision to place his mark there.

For our purposes, however, the test should lend itself to a variety of responses and should be of sufficient interest to hold *S*'s attention. At the same time it must be short and easily understood, since *S* has to fill the blank out five times while doing the experiment. The selection of the Allport-Vernon Scale of Values, Part 2, seemed to satisfy all these requirements. It should be remembered that the test was not selected to enable us to predict regarding different values but only to serve as an operational device to measure changes in scores which are assumed to represent an overt expression of the *S*'s personal constructs.

Procedure. There were three contacts with each *S*: a preliminary administration of the A-V Scale, the interaction periods, and a follow-up sociometric choice contact.[3]

The *S*s were asked in class if they would like to volunteer for an experiment having to do with relationships between people. Those volunteering were given a copy of the A-V Scale and a schedule blank on which they were to indicate the hours they were free to come for the second part of the experiment. The *S*s completed these blanks either in the classroom or between classes and returned them to their instructors.

Same-sex groups of four were formed with these considerations: all four must be able to come at the same time and each one of the *S*s had to have an A-V protocol within a certain limited, arbitrary range of similarity to the score of each of his partners.

Schedules for the second session were then sent out to the *S*s. Each *S* was paired with two other people; e.g., Subject A was paired first with Subject B and second with Subject C, while Subject D was paired first with Subject C and second with Subject B, etc.

Since a discussion of the counterbalanced controls will follow, we will describe the rest of the procedure with no concern for the controls introduced.

When the *S*s first arrived, from one to two weeks after the first session, two of them were placed in a room together and introduced to one another. They were then given copies of the A-V Scale and asked to fill them in as they thought the other person had filled his in. When they were finished, they were separated and each

[3] Sociometric choice data may be found in the original dissertation. In general, sociometric choice was not found to be related to any of the other variables.

given a set of instructions which was designed to create the independent variable of focus of attention. Essentially they were told that the purpose of the experiment was to find our how much they could find out about the other person, *so they should pay attention to him*; or that the purpose was to find out how much information about themselves they were able to communicate to the other person, *so they should pay attention to themselves and what they were saying*. The topic that they were to talk about during the interaction period was also given to the *S*s.

After each *S* had been given his set of instructions, the two were again brought together in the room and told that they could start. After five minutes of interaction, they were interrupted and asked to fill out the A-V again as they thought the other person had filled his out. Each *S* was then introduced to his second partner and the same procedure followed as before.

Controls. Since the projection measure is in terms of the similarity between an *S*'s prediction of his partner's responses and his own responses, the degree of similarity between the partner's own responses and *S*'s own responses might be a possible source of contamination. The possible effect was partially controlled, within the limits placed on the design by having four like-sex *S*s serve at the same experimental time, by matching each pair of *S*s in the degree to which their own A-V protocols were the same. Of the 104 total matchings all but two fell in the range from 7 to 16; 75 of the matchings were in the range from 8 to 11. The maximum range, as the A-V Scale, Part 2, was scored, would have been from zero to 30. .

The possible effect of the topic discussed in the interaction situation was handled to some degree by using two topics. The topics were: (*a*) a vacation that the two people would like to take together, and (*b*) Ohio State University, its advantages and disadvantages, etc. The order of presentation of the topics was systematically varied.

Each *S* was his own control in the variable of focus of attention, that is, each *S* was at one time asked to focus his attention upon himself and at the other time asked to focus his attention upon the other person. The order of focusing attention was also systematically varied.

Since each of the persons in the interaction situation was given a set of instructions and each of them served as *S*s, the nature of the instructions to the "other" person had to be considered. For this reason approximately half of the people, when they were instructed to focus attention upon themselves, interacted with people who were similarly instructed; the other half interacted with *S*s who had been instructed to focus their attention upon the other person. The same kind of control was exercised when *S*s were instructed to focus their attention upon the other person.

In order to control the *S*s' knowledge of the experimental purpose, their instructions contained the notion that the purpose of the experiment was that mentioned in the focus-of-attention instructions. In order to stimulate the discussion between *S*s and to provide some diversionary purpose to the experiment, the *S*s were also told that a recording was being made and that the study was also being conducted with high school students in order to compare how stimulating the conversation of college students is as compared with high school students.

Data analysis. In the last section the procedure in the research design was described. In brief, each *S*: (*a*) took the second part of the Allport-Vernon Scale of Values; (*b*) filled in the scale trying to predict how another person filled his in; (*c*) received a set of instructions concerning his focus of attention while talking with his partner; (*d*) engaged in a five-minute talk with the partner; (*e*) again predicted how his partner filled his in; and, (*f*) repeated the sequence, except for the first step, with another partner. For each *S*, then, we have his own A-V protocol (1-on-1), his first prediction of his first partner, (1-on-2A), his second prediction of his first partner, (1-on-2B), and his two predictions of his second partner, (1-on-3A and 1-on-3B). This makes a total of five protocols for each *S*.

The scoring of the protocols was facilitated by punching each protocol on an IBM card. Since the scoring of accuracy of prediction and assimilative projection entailed the counting of identical marks on two protocols, the score could be easily obtained by noting visually the number of duplications of punched holes.

The A-V Scale, Part 2, consists of 15 questions, each followed by four possible answers. The *S* is asked to fill in the scale, indicating his order of preference among the four possible answers by placing a number, 1 through 4, next to the answers, 1 denoting the highest choice or most preferred answer, 2 the next choice, and so forth down to 4, the last choice or least preferred answer. In scoring the protocols the first and last choice, or most and least preferred, of the responses to each question were used. This method gave an adequate range, from zero to 30, for each test.

RESULTS

Assimilative projection. Hypothesis I stated that focusing attention upon himself should increase the degree to which *S* projects his own constructs upon the other person. This leads to the prediction that *S*'s projection score, i.e., the correspondence between his prediction of his partner's responses and his own responses, will be higher after he interacts with his partner, with attention upon himself, than it was before this interaction. The results confirm the prediction. The mean of the projection scores before interaction was 12.42, and after interaction 13.50. The difference of 1.08 (SD diff = .505) yields a t value of 2.13 ($p = .02$).

Hypothesis II is the crucial test of the differentiating effect of the focus-of-attention variable. Attention on self should increase the projection scores more than attention on the others. This hypothesis predicts that when change scores are computed for the self-and-other situations, the change should be a greater increase for attention on self than for attention on others. The mean increase in the self-situa-

tion was +1.08, in the other-situation, −1.10. The difference of 2.18 (*SD* diff. = .700) yields a *t* value of 3.11 ($p < .001$).

Accuracy of prediction. Hastorf and Bender (4) have pointed out that accuracy of prediction may be influenced by projection which just happens to be accurate. For this reason we canceled out of the accuracy score those scores which were due to projection.

A gross accuracy score was obtained by comparing the prediction protocol (1-on-2A) with the other person's protocol (2-on-2). Next, the identity of three protocols, the prediction protocol (1-on-2A), the other person's protocol (2-on-2), and the self protocol (1-on-1) was determined. The number of identities of all three protocols was then subtracted from the gross accuracy score to obtain the refined score. This score, then, represented the number of times the *S*s predicted the other person accurately, with the number of times he predicted accurately *but also projected* taken out of consideration.[4]

Hypothesis III stated that focusing attention upon his partner should increase the degree to which *S* accurately predicts his partner. The prediction, then, is that *S*'s refined accuracy score will be higher after he interacts with his partner, with attention on his partner, than it was before this interaction.

The results again confirm the prediction. The mean of the refined accuracy scores before interaction was 4.56, and after interaction 5.39. The difference of .83 (*SD* diff. = .317) yields a *t* value of 2.61 ($p < .01$).

Hypothesis IV stated that attention on others should increase the refined accuracy scores more than attention on self. This hypothesis predicts that when change scores are computed for the other- and self-situations, the change should be a greater increase for attention on other than for attention on self. The mean increase in the other-situation was +.827, in the self-situation, −.327. The difference of 1.15 (*SD* diff. = .475) results in a *t* value of 2.43 ($p < .01$).[5]

DISCUSSION

The positive findings in support of Hypotheses I and II indicate that assimilative projection is a function of paying attention to oneself. We have considered assimilative projection as the result of construing another person while bringing that person within the range of convenience of constructs about oneself. In our way of thinking, construing another person is necessitated by talking with, talking about, or thinking about that person. And the construing is of a projective kind, we hold, when the person is focusing his attention upon himself.

The finding that there is an increase in accuracy of prediction in the attention-toward-other situation (Hypotheses III and IV) adds to our proposition; for it indicated that the opposite of incorporation, differentiation, can be facilitated by focusing attention upon the person who is to be differentiated. And further, this differentiation leads to increase in accuracy of prediction. We reason that constructs are more and more applicable to the person and thus more accurate when elements in those constructs have been behaviors, or events, directly associated with that person. It is with attention upon the other person, of course, that we are able to perceive those elements. Attention, then, serves not only to eliminate references to other objects in the situation, of which the self is the most important, but also to facilitate the incorporation into constructs about the person of elements which are appropriate and applicable to that person.

Since the Würzberg school, the function of attention in furnishing the key to action and thought has been known. Here we have used attention as a determining tendency in the projective perceptions of other people. What we seem to have demonstrated at a very

[4] Whether or not the projection scores should also be controlled for accuracy in the same sense that accuracy scores are controlled for projection should be considered. At this time the present author feels that the more meaningful operation is to include all the projection score regardless of whether or not it is accurate. A "refined" projection score was determined, however, and for those interested, the *t* for the increase in "refined" projection in the "aware-of-self" situation was significant between the .05 and the .01 levels. The difference in the increase between the "self" and "other" situations was significant beyond the .001 level of significance.

[5] Dr. Lee J. Cronbach, in a personal communication, dated January 17, 1955, has noted that the increased projection and the increased refined accuracy scores are not independent, since they are based on the same instrument. An increase in refined accuracy can be predicted in situations in which there is decreased projection due to the score linkage. For this reason Hypotheses I and III, and Hypotheses II and IV have not been given distinct experimental tests.

general level is that construing another is in terms of assimilative projection in situations in which incorporation of the other person into already existing constructs about the self is facilitated by focusing attention upon the self. Construing another person is more in terms of an accurate prediction of that person in situations in which discrimination of self from other is facilitated by focusing attention upon the other person.

Implications for clinical psychology. An increasing amount of research has been directed at solving some of the problems in the therapeutic relationship. Though the notions to be tested in our problem did not spring from a clinical analysis of the therapeutic situation, the contribution which this study can make to our understanding of therapy should be fairly direct. In the first place, the experimental situation of having two people sit across from one another and talk to one another is almost a duplicate of most therapeutic relationships. More important, however, the idea of accurately predicting, or better, accurately construing another person's constructs, is the explicit goal of therapy in the psychology of personal constructs. It may be possible to help a person reorganize his construct system without knowing anything about it; but we certainly assume that a therapist can do a much more reliable job of it if he first can adequately construe the other person and his constructs.

From our study we can suggest that we have some data to support the notion that the therapist, if he wants to understand his patient, should pay attention to him. This may sound prosaic; at the same time we seem to have demonstrated the result of paying attention to oneself or to one's own ideas, as many therapists seem to do. Initial conceptions of the patient may be in the direction of increased projection and decreased accuracy.

To generalize even further, differentiating the patient from others, and thus from oneself, is necessary to the formulation of hypotheses about the patient which are most applicable to him; and the way to do this is to focus one's attention upon him and try to learn important things about him rather than to be concerned with how one is behaving as a therapist. It is in this manner that hypotheses about him have *him* as an important referent rather than the therapist.

We may also be able to make a comment on the training of therapists if we note the brevity of the instructions to the *S*s and the difference which the instructions made "within each subject" in their subsequent conceptualization of the other person. This may mean that therapeutic techniques with elaborate systems of thought that are subsequently taught to the patient facilitate mainly a projection of the therapist's notions about the world. A naive therapist, told only to listen carefully to the patient and to pay attention to what he is saying, may have more accurate conceptualizations about *that* patient than the more "schooled" student.

Fiedler (2, 3) has found that a good therapeutic relationship is one in which the therapist's assimilative projection onto the patient is high. This projection, according to Fiedler, establishes a relationship in which the therapist sees the patient as a person with whom he can identify, and whom he can therefore understand, like, and respect. The present results would seem to indicate that assuming an "understanding" attitude, i.e., paying attention to the other person, tends to decrease assimilative projection. It becomes evident, therefore, that further experimentation to resolve or clarify this apparent contradiction is needed.

SUMMARY

Fifty-two college freshman and sophomore students in the beginning psychology class at Ohio State University, who had volunteered to participate, completed the Allport-Vernon Scale of Values, Part 2. Two to three weeks later they each met with two other *S*s and discussed a topic for five minutes with each of them. Before and after each discussion they filled in the A-V scale as they thought their partners had filled their's in. Before one of the discussions *S* was instructed to focus his attention upon himself; before the other discussion he was asked to focus attention upon his partner.

The results were conceptualized under the proposition that change in *S*s' predictions of their partners was in the direction of greater similarity to the predictor's protocol—increased assimilative projection—when the predictor focused his attention upon himself. Change was in the direction of greater

similarity to the predicted person's protocol —increased accuracy—when the predictor focused attention upon the other person.

Discussion of the findings suggested that assimilative projection occurs in situations in which incorporation of the other person into already existing constructs about the self is facilitated by focusing attention upon the self. Accurate prediction of the other person takes place when differentiation of self from other is facilitated by focusing attention upon the other person.

The implications of this study for clinical psychology were discussed.

REFERENCES

1. BIERI, J. Changes in interpersonal perceptions following social interaction. *J. abnorm. soc. Psychol.*, 1953, **48**, 61–66.
2. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101–107.
3. FIEDLER, F. E., & SENIOR, KATE. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. *J. abnorm. soc. Psychol.*, 1952, **47**, 446–453.
4. HASTORF, A. H., & BENDER, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574–576.
5. KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

Received January 22, 1955.

CONCEPT FORMATION AND IMPRESSIONS OF PERSONALITY[1]

EUGENE S. GOLLIN
Institute of Child Welfare, University of Minnesota
AND SHELDON ROSENBERG
University of Minnesota

IN AN earlier study it was found that marked individual differences existed in the formation of impressions of personality (3). It was suggested that these differences reflected variations in subserving perceptual-cognitive organizing processes, and that these processes might represent general personological trends.

In the impression-formation situation *S*s were shown a short motion picture which portrayed a young woman, the Star, in a series of discrete and diverse behavioral activities. There were two major behavioral themes in the motion picture, promiscuity and kindness. In response to the request to present an impression of the personality of the Star, some *S*s characterized her in terms of both major behavioral themes and attempted to account for the presence of these diverse behaviors by an interrelational statement. Other *S*s characterized the Star by enumerating the major behavioral themes without in any way interrelating them. A third group of *S*s retained only one of the major behavioral themes in their characterization of the Star; for these *S*s the Star was either a nice, kind person, or an immoral person, but not both.

The *S*s who form impressions of this last kind fail to articulate the materials presented to them for judgment. One particular aspect of observed behavior dominates their written impressions. For discussion purposes we will refer to these impressions as simplified.

Impressions of the first kind, where interrelation is attempted, will be referred to as related impressions, and impressions which retain both themes but contain no interrelational statement will be called aggregated impressions.

In both related and aggregated impressions the several essential molar features of the judgmental field are adequately articulated. In the related impression something is added to the articulation. The *S*s forming related impressions go beyond the perceptually given in an attempt to fulfill the request to form an impression of personality; *S*s forming aggregated impressions do not give evidence of going beyond the perceptually given.

Correlates of the three impression-formation procedures within the framework of the impression-formation situation itself are reported elsewhere (3). Our present interest is to determine whether the organizational tendencies revealed in the impression-formation task carry over into judgmental situations with vastly different content.

Such a judgmental situation is provided by Rokeach's interrelations task. Rokeach presented *S*s with 10 terms; *Buddhism*, *Capitalism*, *Catholicism*, *Christianity*, *Communism*, *Democracy*, *Fascism*, *Judaism*, *Protestantism*, and *Socialism*. Five are religious terms and five are political-economic terms. After his *S*s had defined these terms, Rokeach had them write a paragraph indicating how they thought the terms might be interrelated (5).

Our interest in this task is that it provides a situation quite different from the motion picture as far as content is concerned while at the same time permitting individual differences in field organization to emerge. Our instructional paragraph to *S*s was intended to be less suggestive than that of Rokeach and we utilized a different set of jugmental criteria in categorizing the protocols. We were interested in utilizing the Rokeach task to divide our *S*s into two groups, those who would organize the 10 terms by means of a hierarchic concept and those who would not accomplish such organization. By hierarchic concept is meant a verbal statement which clearly articulates the major subgroups and organizes them into a category such as "social institutions," "beliefs," "methods of group control," etc.

[1] The data utilized in this study were gathered at Brooklyn College. The authors wish to express their gratitude to A. Gorvine, M. Newman, and O. Schlesinger and their students for their cooperation. A report of this investigation was read at the 1954 meetings of the American Psychological Association.

Cognitive theorists account for "styles" of response in terms of representational processes (4, 6). Examples of these processes or mediational dynamisms are Frenkel-Brunswik's "tolerance of ambiguity" (1), and Goldstein and Scheerer's "abstract and concrete capacity levels" (2). As definition of central dynamisms of this sort proceeds, their utility for theory and experiment increases. This investigation is aimed at providing further specification of postulated central representational systems.

If such systems are operative, then *S*s who fulfill the impression of personality task by articulating and relating aspects of the field should be more likely to form hierarchic concepts in the Rokeach test than others. The *S*s who form simplified impressions should be least likely to form hierarchic concepts in the Rokeach task; *S*s forming aggregated impressions are likely to fall between the other two groups since in the impression-formation task they articulate the several essential molar features of the field but fail to present an interrelating statement.

One further expectation is entertained. In the earlier impression-formation study it was found that the manner in which impression-formation proceeded was related to social-distance scale judgments of the Star. The *S*s who formed simplified impressions were significantly more extreme in their social-distance judgments than either those who formed aggregated or related impressions. Since it is postulated that similar process activities subserve both the simplified impression and the lack of hierarchic concept formation, it is expected that *S*s who do not organize the Rokeach task by means of a hierarchic concept will tend to be more extreme than other *S*s in social-distance judgments of the Star. This expectation is supported by the work of Frenkel-Brunswik who reports that individuals who reveal cognitive intolerance of the coexistence of disparate features in the same object or person tend to make black-white judgments, to arrive at premature closure with regard to valuative aspects of situations, and to accept or reject people in over-all, unambiguous, and unqualified fashions (1). To form a hierarchic concept is not only to tolerate coexistences, but to go beyond that which is given, and to order disparate features by a particular kind of organization. This organization embodies many of the characteristics which Goldstein and Scheerer ascribe to an "abstract attitude." It is quite different from overcommitment to particular features of the stimulus field.

METHOD

The *S*s were 121 evening session undergraduate students. They were shown the motion picture and asked to write their impressions of the personality of the Star. They then responded to the following social-distance scale:

"The questions listed below all concern the person you just wrote about. Using the following scale, answer *in order* each of the questions by placing the number which represents your choice in the appropriate space.

YES	1
PROBABLY	2
DON'T KNOW	3
PROBABLY NOT	4
NO	5

Would you object to her working in the school cafeteria?_____.
Would you object to her living in your neighborhood?_____.
Would you entrust the handling of money in a business to her?_____.
Would you object if she were a personal friend of your sister, mother, or wife?_____.
Would you object to having her in your classes?_____.
Would you object to her working with little children?_____.
Would you object if a close relative of yours wanted to marry her?_____.
Would you object to having to work with her?_____.
Would you want her to join your social group?_____."

On another evening these *S*s were given the Rokeach task. The following instructional paragraph was presented:

"Now that you have defined the terms here is what we would like you to do next. In the blank space provided below we would like you to write a paragraph stating whether or not you feel that all of these terms are interrelated. If you feel that all of these terms are interrelated *describe the way in which you think they are interrelated*. If you feel that these terms are not all interrelated *then just write about those terms which you think are interrelated*. Do not worry about how well organized your paragraph is because that is not important for the purposes of this study. You will have about 5 minutes to do this in."

The order of presentation of the impression forming test and the Rokeach test were roughly counterbalanced.

RESULTS

All the written materials in the two test situations were categorized independently by the authors. Judgmental agreement in both cases exceeded 89 per cent.

In the impression forming task 18 per cent of the sample wrote impressions categorized as related, 24 per cent wrote impressions categorized as aggregated, and the remaining 58 per cent wrote impressions which were classified as simplified.

In the Rokeach task 38 per cent of the sample produced hierarchic responses while the remaining 62 per cent did not.

A 2 × 3 chi-square table was constructed to determine if a statistically significant relationship existed between judged performance in the impression forming task and judged performance in the concept forming task. Table 1 contains the respective distributions. Of the *Ss* forming related impressions, 68 per cent also formed hierarchic concepts. Only 36 per cent of those writing aggregated impressions and only 30 per cent of those writing simplified impressions presented concepts of a hierarchic nature. The statistical comparison yielded a significant chi-square value (10.68, $df = 2$, $p < .01$).

To obtain further information concerning the relationship indicated by the over-all chi-square value, a series of fourfold tables were constructed and chi squares, corrected for continuity, were run. The *Ss* in the related impression category differ significantly from both other groups with respect to concept formation, but there is no statistically significant difference between the aggregated and simplified groups.

The results of the social-distance analysis indicated that *Ss* who do not form hierarchic concepts are markedly more extreme in their social distance scale responses than *Ss* who do form such concepts. This is exemplified by the fact that 39 per cent of the non-concept-formers in this study have social distance scale mean scores falling in the middle two quartiles of the distribution as compared with 73 per cent for the concept formers. In each of the two extreme quartiles there are approximately 30 per cent of the non-concept-formers as compared with only 13 per cent of those who form hierarchic concepts. The means and *SD*'s of the two groups are presented in Table 2. While the social-distance scale means of the two groups are virtually identical, the variance of the concept-forming group is significantly smaller than that of the non-concept-forming group. The hypotheses that the latter group would tend to be more extreme in their social-distance scale responses is supported.

TABLE 1

DISTRIBUTION OF CONCEPT-FORMERS AND NON-CONCEPT-FORMERS AMONG IMPRESSION-FORMATION CATEGORIES*

Impression Categories	Concept-Formers	Non-concept-Formers	*N*
Related	15	7	22
Aggregated	10	18	28
Simplified	21	50	71
N	46	75	121

* $\chi^2 = 10.68$, $p < .01$.

TABLE 2

SOCIAL-DISTANCE SCALE MEANS AND STANDARD DEVIATIONS

Group	*N*	Mean	*s*	s^2	*F*
Concept-formers	46	20.32	5.58	31.13	2.14*
Non-concept-formers	75	20.41	8.16	66.58	

* For $p < .01$, $F > 1.90$.

DISCUSSION

The findings lend support to the hypothesis of generality of cognitive style. A relationship is apparent between organizing tendencies in impression formation and concept formation.

In an earlier study it was found that subjects who formed Simplified impressions were extreme in their social-distance scale selections (3). In the present investigation a similar relationship was obtained between concept formation and social-distance scale response. Subjects who do not form hierarchic concepts tend to exceed concept formers in the extent to which they accept or reject the Star as measured by the social-distance scale described above. Thus, *Ss* who apparently utilize similar cognitive techniques achieve scores at the opposite ends of a continuum. This investigation has emphasized structural features of behavior. It is likely that analysis of content would also provide worthwhile information. However, it should be noted that emphasis on structural features provides an important means for approaching questions of consistency in behavior.

There are a number of further research questions which need exploration. What are the sources of these response tendencies? Some

evidence regarding this question is presently available. It has been found that certain of the response characteristics manifested in adult samples arise developmentally; that is to say, lack of articulated response, enumeration of events, and going beyond the perceptually given are, in turn, more prominent as the chronological ages of *Ss* increase from eight to seventeen years.[2] It remains to be determined whether these characteristics of performance relate to such factors as intelligence and social experience.

Still another research question is concerned with the crystallization of the several response patterns at the adult level, and the degree of facility with which adults can mobilize different patterns as environmental demand shifts.

Also of considerable interest, both from a developmental vantage point and with adults, is whether the operation of the inferred central dynamisms influences various kinds of learning; that is, are there consistencies between perceptual-cognitive behavior and learning behavior?

SUMMARY

This investigation was designed to discover whether individual differences in the way in which impressions of personality are formed are related to performance in a concept-formation task. It was hypothesized that *Ss* who articulate and relate the several essential molar features of the impression-formation material are more likely to form hierarchic concepts than *Ss* who articulate the field without relating the various aspects, or *Ss* who fail to differentiate between major field characteristics. It was also hypothesized that *Ss* who fail to differentiate major field characteristics are less likely to form hierarchic concepts than *Ss* who form impressions in which there is articulation but no relational attempt. The first hypothesis is supported by the findings.

One other hypothesis was tested: that *Ss* who form hierarchic concepts are less extreme than non-concept-formers in their social-distance scale responses. The results also support this contention.

[2] A report on this developmental study is in preparation.

REFERENCES

1. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
2. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
3. GOLLIN, E. S. Forming impressions of personality. *J. Pers.*, 1954, **23**, 65–76.
4. LEEPER, R. W. What contributions might cognitive learning theory make to our understanding of personality? *J. Pers.*, 1953, **22**, 32–40.
5. ROKEACH, M. A method for studying individual differences in "narrow-mindedness." *J. Pers.*, 1951, **20**, 219–233.
6. SCHEERER, M. Personality functioning and cognitive psychology. *J. Pers.*, 1953, **22**, 1–16.

Received January 11, 1955.

"CONCEPTUAL LEVEL" VS. "CONCEPTUAL AREA" ANALYSIS OF OBJECT-SORTING BEHAVIOR OF SCHIZOPHRENIC AND NONPSYCHIATRIC GROUPS

LAURENCE S. McGAUGHRAN[1]
University of Houston
AND LOUIS J. MORAN[2]
VA Hospital, Houston, Texas

It frequently has been held that schizophrenia is accompanied by a marked loss of the ability to conceptualize on an abstract level, and by an increased tendency toward the use of concrete forms of conceptualization. Others have suggested that the form and content of schizophrenic thought reflect not so much a loss in basic abstractive ability as they do a lessening in the ability, or the intent, to communicate freely and consistently with others.

Those who have regarded schizophrenia as a shift in level of conceptual ability (or "attitude") have usually been identified with the theoretical orientations of Goldstein or Vigotsky; investigators within this group include Goldstein, Bolles, Storch, Vigotsky, Kasanin, Hanfmann, Zeigarnik, and Rapaport. Characteristic of this position of a qualitative distinction of conceptual levels is Goldstein's recent statement:

> Even in its simplest form, however, abstraction is separate in principle from concrete behavior. There is no gradual transition from the one to the other. The assumption of an attitude toward the abstract is not more complex merely through the addition of a new factor of determination; it is a totally different activity of the organism (6, p. 60).

Elsewhere, Goldstein specifically describes the concreteness of schizophrenic thinking as follows: "We can understand schizophrenic (symptoms) as actions which are commensurate with the changed world of the schizophrenic, a world which is pathologically concrete and void of abstract interpretation" (5, p. 34).

The terms, *concrete* and *abstract*, generally have been used by this group to differentiate between the conceptual behavior of normal adults and that of such deviant groups as the brain-damaged, schizophrenics, and children. According to their position, an abstract attitude may be achieved only by normal adults; individuals outside of this class are characteristically concrete in their conceptualization. While some of the investigators (Hanfmann and Kasanin [8], Bolles [1], Rapaport [13]) have suggested breaking down the abstract-concrete dichotomy by isolating additional levels (e.g., "partial identity," "co-functionality"), these levels have been re-incorporated in the analysis of data to coincide with the rationale for one or the other poles of the dichotomy. Thus, the end effect of this type of analysis is to treat all concepts as either abstract or deviantly nonabstract.

The views of Sullivan and of Cameron are typical of those who feel that schizophrenia is primarily a disturbance in social communication or "consensual validation" rather than an actual loss of abstractive ability. Sullivan writes: "The peculiarities of language behavior in the schizophrenic arise from his extreme need of a feeling of personal security (14, p. 7). . . . Language operations as thought are profoundly different, quite fundamentally different from language operations as communication and as pure mechanisms used in dealing with others; the more completely one becomes self-centered the more utterly he becomes cut off from integrations with other more or less real people, and the more utterly novel, perfectly magical, and wholly individual become the symbols which he uses as if they were language (14, p. 9). . . . The schizophrenic's speech shows characteristic peculiarities because of recurrent severe disturbances in his relationships with other people and the result is a confusion of the critical faculties concerning the structure of spoken and written language" (14, p. 15).

Similarly, Cameron (3) describes the schizophrenic as "disarticulated" from the social group, as defective in role-taking skills, and unable to anticipate or share the attitudes of

[1] Employed part time at the Houston Veterans Administration Hospital, where this study was conducted.
[2] Now at the University of Texas.

others. He (2) has concluded from his experimental work that peculiarities of the schizophrenic's speech and thought result primarily from disturbances in self-monitoring and communication rather than from an impairment in the basic ability to generalize, or abstract. A somewhat similar position is adopted by Hunt and Cofer (9).

The issue, then, is whether schizophrenic thought reflects a marked loss of ability to conceptualize on an abstract level and an increased tendency toward the use of concrete forms of conceptualization, or whether it can best be characterized as a disorder of social communication skills, which does not necessarily involve impairment in abstractive ability. Experimental attempts to resolve this issue generally have been nonquantitative studies, or have left uncontrolled such possible contributing factors as age, educational level, test-intelligence level, and status of hospitalization. The present study is a quantitative investigation of differences in conceptualization between matched pairs of schizophrenic and nonpsychiatric patients with these factors experimentally controlled.

METHOD

Subjects

The samples consist of a group of 37 male patients diagnosed as chronic paranoid schizophrenics and a group of 40 male patients who were undergoing treatment for or recuperating from nonpsychiatric illnesses in a VA hospital. The schizophrenic group was successfully matched by pair with 37 of the nonpsychiatric group for estimated test intelligence,[1] educational level, and age. None of the group mean differences on these variables approached significance. The matched pairs are used in this study for analysis of mean differences, and the total groups for an analysis of differences in distributions of scores.

The Sorting Task and Definitions

The experimental task used to investigate differences in conceptualization between the groups was the Rapaport modification of the Goldstein-Gelb-Weigl Object-Sorting Test. This task was administered individually to each of the *Ss* in accordance with the procedure described by Rapaport (13). In brief, this technique involves phases of "active" and "compliant" sorting of 33 objects familiar in everyday experience. The active phase consists of seven tasks requiring *S* to group with some particular object the others which "belong with it." The compliant phase consists of 12 tasks requiring *S* to identify the basis for the grouping of a number of objects arranged by *E*.

Two methods of scoring were employed; responses were scored first for "conceptual level" and then for "conceptual area."

Conceptual-level scoring method. This method corresponded with Rapaport's system for scoring "conceptual level." Five basic measures were employed: abstract, functional, concrete, failures, and pathological indicators.

1. *Abstract*—Rapaport defines the abstract-conceptual level as one in which the "essential abstract-conceptual common content (is expressed) in a generic term" (13, p. 403). Since the judgment of "essentiality" without well-established norms could become subjective, the basic measure for this category was scored directly from Rapaport's list of acceptable responses for each test item. (Sample: "tools," "whiteness.")

2. *Functional*—Rapaport defines the functional level as one in which objects are grouped together "because of a common function performed with or on them by human beings. . . ." (13, p. 403). (Sample: "used in smoking.")

3. *Concrete*—Rapaport defines the concrete level as one in which the grouping is based upon "inessential detail" (13, p. 403). Here, again, to avoid the possibility of subjectivity in judgment of essentiality, items were scored as "concrete" if they were not scorable in any other category. (Sample: "If you wrote on paper, you might need an eraser.")

4. *Failures*—Rapaport scored "failure" if no sorting was made on the active phase or no definition offered on the compliant phase.

5. *Pathological indicators*—This group includes a number of special categories employed by Rapaport which were, in general, considered to indicate conceptual pathology. The separate terms are as follows: syncretistic, fabulated, symbolic, chain, split-narrows, and false. They were scored together under this grouping because responses within the separate categories usually occur too infrequently for statistical analysis and because responses of this nature frequently contain elements of several of the separate types. (Sample: "All have weight, value, and are used by man.")

Two other measures derived from the basic abstract category are as follows:

6. *Abstract II*—this includes the responses scored in the basic abstract category plus additional responses which were conceptual classifications but were not scored as abstract definitions by Rapaport. (Sample: "Both are composed of sugar.")

7. *Abstract III*—this includes all of the responses in Abstract II plus additional concepts which are classifiable as pathological but are conceptual classifications of a relatively high order of abstraction (Sample: "All are forms of energy.")

Conceptual area scoring method. The basis for scoring conceptual areas has been described in detail in previous studies (4, 10). However, since the present study involves somewhat different object-sorting and scoring procedures and a different population of *Ss*, further explanation seems appropriate. The attempt to work out differences in areas of conceptualization has involved the use of four areas, which are represented as

[1] Test intelligence was estimated on the basis of vocabulary scores on the Wechsler-Bellevue Scale. Justification for this procedure is based upon the reported correlation of .85 between vocabulary and full scale scores (15).

quadrants formed through the intersection of two variables of conceptualization. The first variable is scored in terms of the relative "publicness" or "privateness" of each separate conceptual definition, and may be termed the public-private variable. A relatively high degree of publicness of a concept implies that the principle underlying the grouping is shared and freely communicated by the majority of persons within the same culture. (For example, the boundaries for a concept of "triangularity" are culturally denotable; the conceptual boundaries for "attractiveness" would be considerably less so; thus, the exact meaning intended by the latter term is not easily communicated.)

The second variable is scored in terms of the number of common attributes which are used, or implied, in each conceptual definition. Thus, a concept based upon a single attribute common to all objects within the conceptual grouping would be maximally "open" (e.g., a group of various objects conceptualized as belonging within the same class because all are red in color). On the other hand, a concept binding together all perceptible attributes, all of the "degrees of freedom," of a group of objects within a conceptual grouping would be maximally closed (e.g., a concept based upon the absolute identity of all of the grouped objects). It can be seen that the more open a concept the more abstract or higher order is the principle of conceptual classification involved. This shall be referred to subsequently as the open-closed variable.

Thus, if these two variables, with their respective poles of public-private and closed-open, are visualized as intersecting in space, the resulting quadrants may be conceived to represent the following four conceptual areas: closed-public, open-public, open-private, and closed-private. These four conceptual areas are the basic scoring categories. Instead of attempting a full description[4] of all of the subcategories within these areas, we shall cite examples of typical conceptual groupings within each area.

1. *Closed-public area*

a. Absolute identity—(2 sugar cubes; basis: "exactly alike").

b. Species identity—(real and toy screwdrivers; basis: "screwdrivers").

c. Reciprocal co-functionality—(hammer, nail; basis: "need hammer to drive nail").

d. Induced partial identity—(2 sugar cubes, file card; basis: "white and rectangular").

2. *Open-public area*

a. Unitary abstraction—(red circle, red ball, red eraser; basis: "they're red").

b. Nominal superordination—(several toys; basis: "all these are toys").

c. Implicit superordination—(several toys; basis: "all of these you use in play").

3. *Open-private area*

a. Physiognomic—(several objects; basis: "objects that please me").

b. Representational—(matches, cork; basis: "burnt cork for minstrel").

c. Asyndetic—(matches, toy hatchet; basis: "matches from trees and ax reminds me of George Washington, February 22, and the winter when you need heat").

d. Metonymic—(red circle, green square; basis: "mathematical figures").

4. *Closed-private area*

a. Specific mediation—(bell, lock; basis: "put bell on bicycle and lock it with lock").

b. Experiential contiguity—(green square, matches, cigarette; basis: "have on my desk at home").

c. Closed radial—(matches, smoking objects; basis: "need matches to light each of these").

Scores within these four areas are interrelated to derive three additional measures: total public vs. private, total closed vs. open, and an "autistic index" (open-private minus closed-public). Interscorer agreement with the basic measures of conceptual area is relatively high; with a series of 55 random items, there was perfect agreement on 51, or 93 per cent.

Procedure

The object-sorting task was administered to all *S*s by one experimenter and was independently scored by the other. All identifying characteristics of the protocols were masked and random numbers assigned to each *S*. All scoring was completed before the statistical analysis of the data was begun.

Hypotheses

Two forms of analysis of object-sorting data will be attempted. The first of these will involve measurement in terms of *conceptual levels* and is intended to test the general assumption that schizophrenia is accompanied by a loss of abstractive ability or attitude. Here, the major determination is whether each concept is abstract or nonabstract in form. The second form of analysis will involve measurement in terms of *conceptual areas* and is intended to test the general assumption that schizophrenic thinking is accompanied by a loss of social communication (loss of "publicness") in the formation of concepts, without evidence of a corresponding loss in high order of conceptualization (loss of "openness").

The specific hypotheses for the first form of analysis (conceptual level) are as follows. The schizophrenic group will:

1. employ significantly fewer abstract-conceptual definitions than will the nonpsychiatric group;

2. employ significantly more of the following categories of sorting than will the nonpsychiatric group: (*a*) functional, (*b*) concrete, (*c*) failures, (*d*) pathological indicators.

The senior author has previously employed with normal *S*s an analysis of "conceptual areas" as defined here. In this study the prediction was made that a tendency toward the predominant use of closed-public concepts would be "associated with a more general behavioral potential toward 'reality-fixing' or 'stimulus-

[4] A complete table of the subcategories and examples of the scoring of conceptual areas as used in this study has been deposited with the ADI. Order Document No. 4621 from the ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.50 for photoprints or $1.75 for 35 mm. microfilm. Make checks payable to Chief, Photoduplication Service, Library of Congress.

boundness' " (10, p. 188) and that the predominant use of open-private concepts would be "associated with a more general behavioral potential toward 'autism' " (10, p. 188). On the basis of this distinction, the language behavior of representative groups of normals was successfully predicted in a number of language measures which were intended to relate to such specific behavioral potentials as rigidity-fluidity, inhibition-spontaneity, passivity-autonomy in reality-structuration, vigilance-laxness in reality-testing, etc. On the basis of these tentative findings, and under the assumption that the peculiar forms of schizophrenic conceptualization should be more closely associated with autism and reality loss than with reality fixing and stimulus-boundness, the specific hypotheses to be tested in the analysis of conceptual area are as follows. The schizophrenic group will:

3. employ significantly fewer public concepts and significantly more private concepts than will the nonpsychiatric group;

4. employ significantly fewer closed concepts and significantly more open concepts than will the nonpsychiatric group;

5. employ significantly more (*a*) open-private, (*b*) and closed-private concepts, and significantly fewer (*c*) closed-public concepts than will the nonpsychiatric group;

6. score significantly higher on an "autistic index," consisting of open-private minus closed-public concepts, than will the nonpsychiatric group.

RESULTS

The significance of differences in distribution of scores for the schizophrenic and nonpsychiatric groups on the conceptual level measures was estimated initially by chi square. The results of this analysis are shown in Table 1. Since an analysis of score distributions cannot account for the effect of possible correlation between the measures of conceptual level and the control variables, critical ratios of mean differences between matched-pair groups were also computed. The differences in p values yielded by the two forms of statistical analysis clearly indicate that much of the variance is attributable to the effects of the controlled variables.

As Table 1 indicates, neither system of statistical analysis supports a single hypothesis concerning conceptual levels at any value approaching significance. In fact, the results of two of the five basic measures (functional and failures) show a significant (.03 level) difference between the groups in a direction opposite that hypothesized. The results obtained with the basic abstract measure are also in the opposite direction to that hypothesized. However, the analysis of mean difference indicates that the higher abstract scores for the schizophrenic group could easily be attributed to chance.

When both the extent and direction of group difference are formally hypothesized and a one-tailed test of significance is employed, there is some difference of opinion among statisticians concerning interpretation of results which may be statistically significant but in a direction opposite to that hypothesized. Because our results with the functional and failure measures are of this nature, we shall not attempt to discuss nor to interpret the significance of these findings except to observe that they do not support the hypotheses concerning differences in conceptual level between our schizophrenic and nonpsychiatric groups.

A similar analysis was conducted for the conceptual area measures. The results are shown also in Table 1. While all of the results shown in the chi-square column indicate group differences in the predicted direction, the only significant difference is that for index measure associated with autism (open-private minus closed-public). However, when the variance

TABLE 1

SIGNIFICANCE OF DIFFERENCES BETWEEN SCHIZOPHRENIC AND NONPSYCHIATRIC GROUPS ON "CONCEPTUAL LEVEL" AND "CONCEPTUAL AREA" MEASURES

Measure	χ^2 (N = 77)	p*	Mean Scores		σ_{md}	CR (N = 74)	p*
			Schizophrenic	Nonpsychiatric			
Conceptual Level:							
Abstract I	3.67	<.06	7.49	6.89	0.77	0.78	—
Abstract II	3.71	<.06	8.19	7.22	0.85	1.14	—
Abstract III	4.68	<.05	13.22	11.32	0.83	2.29	<.03
Concrete	0.08	—	1.51	1.27	0.28	0.86	—
Functional	2.98	<.09	1.65	2.43	0.34	2.29	<.03
Failures	7.24	<.01	1.89	3.16	0.57	2.23	<.03
Pathological	0.14	—	5.75	4.86	0.87	1.02	—
Conceptual Area:							
Public	0.29	—	11.00	12.35	0.92	1.47	<.07
Open	0.29	—	12.95	11.68	0.81	1.57	<.06
Closed-Public	2.35	—	3.78	5.00	0.71	1.72	<.04
Open-Public	0.01	—	7.22	7.35	0.77	0.17	—
Open-Private	2.21	—	5.73	4.32	0.86	1.64	<.05
Closed-Private	0.56	—	2.35	2.30	0.41	0.12	—
Autistic Index†	3.78	<.03	11.94	9.32	1.41	1.86	<.03

* p values are for one-tailed test of significance except for: (*a*) Abstract I, Functional, and Failure, which are in a direction opposite to that hypothesized, and (*b*) Abstract II, Abstract III, and Open-public, for which no hypothesis was attempted.

† Open-private minus Closed-public, plus 10.

attributable to the controlled factors is accounted for by computing critical ratios of mean differences, five of the seven measures have p values of .07 or less.[5] Mean differences for the other two measures were approximately zero.

Since the controlled variables affected the results appreciably, an additional analysis was conducted with two of them (estimated test intelligence and educational level) so that their effects upon the results of the conceptual measures might be more directly observed. The product-moment correlation coefficient between these two variables was .55 for the schizophrenic group and .62 for the controls. Each group, schizophrenic and nonpsychiatric, was dichotomized into high and low subgroups on the variables of test intelligence and of educational level. Chi squares of the differences in distribution of scores on the conceptual measures for these subgroups were computed. The results are shown in Table 2 for both the conceptual level and the conceptual area categories.

A glance at Table 2 will indicate the extensive relationship of test intelligence and education with conceptual performance in the object sorting task. Eleven of the 14 measures are significantly related with either or both of these factors ($p < .05$). Two of the remaining three measures (concrete and closed-public) are similarly related at the .07 and .06 levels, respectively. Only the index measure assessing tendency toward autism remains as apparently unrelated to either test intelligence or educational level.

Within the schizophrenic group, high intelligence and educational level appear to be positively associated with abstract, public, and open-public definitions, and low intelligence and education with pathological, open-private, and closed-private definitions. Within the nonpsychiatric group, high intelligence and educational level appear to be positively associated with abstract and open definitions and low intelligence and education with concrete, functional, and closed-public definitions, and with failures.

Over all, it seems apparent that the major difference between the schizophrenic subgroups based on intelligence and education is in the relative publicness-privateness of their conceptualization, while the major differences between the nonpsychiatric subgroups is along the open-closed variable.

[5] F ratios to test for homogeneity of variance were computed for all measures in Table 1 for which critical ratios and p values are shown. The results of these tests indicate that the hypothesis of a common population variance is tenable in all cases except for the measure of "failures" shown in Table 1.

TABLE 2

SIGNIFICANCE OF DIFFERENCES IN "CONCEPTUAL LEVEL" AND "CONCEPTUAL AREA" MEASURES FOR SUBGROUPS OF SCHIZOPHRENIC AND NONPSYCHIATRIC SUBJECTS DICHOTOMIZED ON INTELLIGENCE AND EDUCATION

Measure	Schizophrenic (N = 37)				Nonpsychiatric (N = 40)			
	Intelligence High vs. Low		Education High vs. Low		Intelligence High vs. Low		Education High vs. Low	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Conceptual Level:								
Abstract I	11.99†	<.001	6.31†	<.02	1.60†	—	4.91†	<.05
Abstract II	9.92†	<.01	6.31†	<.02	5.23†	<.05	4.82†	<.05
Abstract III	4.66†	<.05	3.24†	<.08	4.91†	<.05	3.56†	<.07
Concrete	3.28	<.08	0.76	—	2.56	—	3.51	<.07
Functional	0.02	—	0.02	—	6.67	<.01	0.82	—
Failure	0.07†	—	0.03	—	0.26	—	5.41	<.02
Pathological	5.82	<.02	4.88	<.05	0.10	—	0.42	—
Conceptual Area:								
Public	7.78†	<.01	6.11†	<.02	1.10†	—	0.47†	—
Open	1.34†	—	2.18†	—	3.64†	<.06	4.82†	<.05
Closed-Public	0.25†	—	1.34	—	3.64	<.06	2.43	—
Open-Public	10.09†	<.01	11.99†	<.001	0.92†	—	2.51†	—
Open-Private	4.56	<.05	1.56	—	0.92	—	1.34	—
Closed-Private	6.15	<.02	2.37	—	0.11	—	0.33	—
Autistic Index	2.18	—	0.25	—	1.60†	—	0.90†	—

† Indicates high scores in the direction of high intelligence or high education subgroups.

DISCUSSION

A recent study by Moran, Moran, and Blake (11) reports the failure of a "conceptual-levels" form of analysis to differentiate the vocabulary performances of schizophrenic and normal *S*s, although a subsequent analysis in terms of other dimensions (12) revealed several striking differences in vocabulary performance.

Similarly, in McGaughran's (10) prediction of language behavior in normal *Ss* on the basis of the predominance of closed-public or open-private conceptualization, it was pointed out that the use of a conceptual-levels form of analysis would not have supported the predictions of group differences because in most cases the conceptual behavior of both experimental groups would have been classified simply as concrete. When one considers also the failure of the conceptual-level form of analysis to support a single prediction in the present study, it becomes apparent that there might be some value in re-examining the assumptions and definitions underlying this approach.

Since a great deal of thoughtful empirical inquiry underlies the study of conceptual levels, however, it would be unwise to draw hasty conclusions on the basis of only the present evidence. Here, we shall remark only briefly about a basic problem of definition which seems to us to deserve much more careful consideration than is feasible in the present analysis.

This problem is that of denoting "abstract" and "concrete" behavior. When Goldstein states that in the abstract attitude, "... the individual thing becomes a mere accidental example or representative of a category" (7, p. 6), it might be inferred that he is simply denoting abstraction as an act of conceptual classification, or, in other words, that abstraction involves the isolation and transduction of a portion of a perceptual field into a conceptual category whose limits transcend the particular features of the immediate situation. If this were the only meaning intended, concrete behavior as the antonym would simply imply absence of this act of abstraction or superordination, and the task of distinguishing the two behaviors would be relatively straightforward. Whether it would be rewarding to make this distinction in terms of the simple presence or absence of abstraction in the individual, however, is a different question, since presumably any organism capable of initiating consistent selective behavior would, by definition, be abstracting. Thus to be meaningful as a variable, conceptual classification would have to be expressed in terms of differences in degree of abstraction, or *order of conceptual classification.*

Actually, of course, neither Goldstein nor those who employ his form of analysis limit themselves to this meaning. In further characterizing the distinction between abstract and concrete behavior, they invoke such varied connotations as "conscious-unconscious," "volitional" or "active" vs. "passive" responses to stimuli, "essentiality vs. non-essentiality," "adequate vs. inadequate," "inner vs. outer claims," "higher vs. lower forms of the unifying process," the use of a "generic (vs. a specific) term," and "well-known" or "common sense" vs. "over-precision." While some of these distinctions would be difficult to define behaviorally, it appears that there are, in addition to the factor of the order of conceptual classification cited above, at least three other potentially measurable variables implied by these terms which are frequently considered to determine (or to overdetermine?) the occurrence of abstract as opposed to concrete behavior. These are: (*a*) essentiality of the abstracted attribute, (*b*) extent of social agreement, and (*c*) verbalization of a generic term. While each of these four variables is frequently cited, or at least implied, as a criterion for an abstract concept, it appears that it is often difficult to apply all of them in the same judgment. For example, Rapaport (13) illustrates concrete definitions with two sample responses to the Wechsler-Bellevue item concerning the conceptual similarity of "dog and lion" as follows: "they both have four legs, hair, etc.," and "they both have cells." Concerning these responses, Rapaport states: "Besides being too limited, a concrete definition is also too inclusive; tables too have legs, humans have hair, and so on. As a matter of fact, the concrete definition in some cases may be pathologically inclusive, as in the response that dogs and lions are similar because they both have cells" (13, p. 149).

In analyzing the first of these responses in terms of the four criteria suggested above, it would seem likely that it might meet the order of conceptual classification criterion (four-leggedness, hairiness) and possibly the criterion of social agreement (depending upon the population), but would presumably fail on essentiality and would clearly fail on the verbalization of a generic term. The second response (of "cells") is a conceptual classification of a high order of abstraction and might well be considered to meet the essentiality criterion (al-

though subjectivity is quite apt to enter into this judgment) but does not specifically employ a generic term, and possibly would not be considered as within "the common treasure of our vocabulary and thinking" (13, p. 149).

The criteria of "essentiality" and "social agreement" are relative terms and fit rather awkwardly into the dichotomy of abstract vs. concrete. On the other hand, they appear to have potential dimensionality and conceivably could be employed as conceptual variables in their own right. The "verbalization of a generic term" as a criterion of abstract behavior appears somewhat more consonant with a dichotomous analysis but may very likely have little to do with the presence or absence of the abstractive process, as such. For example, so frequently the only apparent distinction between abstract and functional concepts is in the language employed by the individual. The distinction between the use of the concept, "kitchen equipment," as opposed to the term, "things you use in a kitchen," seems not to be sufficiently basic to support an inference of a qualitative difference between the conceptual processes underlying the two events. It may well indicate differences in verbal facility, or even possibly verbal intelligence, but there is no apparent difference between the two responses in terms of the order of conceptual classification.

To return to the present study, it might be meaningful to examine briefly the relationship between each of the four criteria of abstract as discussed above and the operations and results of this experiment. The concepts of public-private and open-closed employed as variables in this study seem to correspond reasonably closely with the criteria of social agreement and order of conceptual classification, respectively, but are seen by us to vary somewhat independently. In the course of previous experimental work, the scoring of the relative publicness or privateness of a particular subcategory of conceptual response has been empirically keyed on the basis of the frequency of social usage. That is, subcategories of the conceptual areas have been frequently shifted away from one pole and toward the other on the basis of the frequency with which that type of response occurred. Thus, while the relative possibility of denotation, or "publicness," and the extent of social agreement are not necessarily synonymous, they are seen to bear a close relationship.

The variable of open-closed is scored in terms of the order of abstraction employed in conceptual classification. Here "order of abstraction" is not employed in the Goldstein sense of the presence or absence of abstract behavior; abstraction as a process is assumed to exist to some degree in all perceptual transactions. Rather, order of abstraction is intended to mean the relative amount of "surplus" potentially abstractable items within a conceptual grouping which are not incorporated in the principle of collection for that group; the greater the "surplus," the freer, or higher order, or more abstract, the concept.

Although the criterion of essentiality presumably could be objectified by keying the scoring empirically to extent of social agreement, this does not appear to be a very promising approach. The lack of unilinearity in the meaning of essentiality would inevitably lead to subjective differences and unreliable measurement. For example, we should find it difficult to agree with Rapaport that, in response to the "dog-lion" item on the Wechsler-Bellevue Scale, the reply, "four-legged, mammalian, carnivorous animals covered with hair" is either necessarily "redundant" (as he says) or "functional," or "concrete," nor would we agree that it necessarily deals less with essentiality than the simpler but more abstract response, "animals." It may well demonstrate compulsiveness, or a lack of "common sense," but this is a different question.

As regards the fourth criterion, the use of a generic term, our results with the "functional" definition (in which a generic term is absent) as seen in Tables 1 and 2 indicate that this response was most characteristic of nonpsychiatric *S*s of low intelligence. Thus to a considerable extent this type of response seems to be more a function of verbal intelligence than of behavior pathology.

Our addition of the derived measures Abstract II and Abstract III was intended to revise the basis of the measure to coincide with the first criterion, order of conceptual classification. Under these specific conditions, the conceptualization of the schizophrenic group is seen to be significantly more abstract than that of the controls. One may also see here the close correspondence in basis of classification and in

obtained results between the Abstract III measure and the "open" measure in the conceptual areas system of analysis.

SUMMARY AND CONCLUSIONS

This study examines the experimental issue as to whether schizophrenic thought reflects a marked loss of ability to conceptualize on an abstract level and an increased tendency toward concreteness, or whether it can best be characterized as a disorder in social communication. The samples consisted of 37 pairs of chronic paranoid schizophrenics and nonpsychiatric patients who were successfully matched on age, educational level, and estimated test intelligence.

The experimental task used was the Rapaport modification of the Goldstein-Gelb-Weigl Object-Sorting Test. Two methods of scoring were used. The first method was that developed by Rapaport for scoring "conceptual level" and was used to test several hypotheses consistent with Goldstein's general assumption that schizophrenia is accompanied by a loss of abstractive ability. The second form of analysis involved measurement in terms of four "conceptual areas" which were derived from the interaction of two variables, publicness-privateness, and openness-closedness. This form of analysis was used to test several hypotheses consistent with the assumption of Cameron, Sullivan, and Hunt and Cofer, that schizophrenic thinking is accompanied by a loss of social communication in the formation of concepts.

An analysis of the results of the conceptual level scoring indicated that none of the hypotheses generated from the Goldstein assumption was supported. An analysis of the results of the conceptual area scoring indicated that the direction of group differences was predicted in five of six measures; the p values for the five critical ratios ranged from .03 to .07, using a one-tailed test of significance.

A further analysis of the effects of intelligence and education upon conceptual performance indicated that 11 of the total of 14 measures were significantly related with either or both of these factors at or below the .05 level of confidence. These differences were more pronounced within the schizophrenic group.

On the basis of the results obtained in this study we draw the following conclusions:

1. Within the limits of the operations of this experiment, our schizophrenic group demonstrated a loss of social communication without evidence of impairment in abstractive ability.

2. Differences in conceptualization in the object-sorting task were clearly more closely associated with estimated test intelligence and education than with the presence or absence of schizophrenia. This strongly suggests that these control factors cannot safely be disregarded in experimentation in the conceptualization area.

REFERENCES

1. BOLLES, MARJORIE. The basis of pertinence. *Arch. Psychol.*, 1937, No. 212.
2. CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
3. CAMERON, N. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 50–64.
4. CLEMENTS, S. D. An exploratory study of behavioral correlates of personal constructs: I. Relationships between personal constructs and object sorting behavior. Unpublished master's thesis, Univer. of Houston, 1954.
5. GOLDSTEIN, K. Methodological approach to the study of schizophrenic disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 17–40.
6. GOLDSTEIN, K. *Human nature in the light of psychopathology*. Cambridge: Harvard Univer. Press, 1947.
7. GOLDSTEIN, K. *Language and language disturbances*. New York: Grune & Stratton, 1948.
8. HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Mongr.*, 1942, No. 67.
9. HUNT, J. MCV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
10. MCGAUGHRAN, L. S. Predicting language behavior from object sorting. *J. abnorm. soc. Psychol.*, 1954, **49**, 183–195.
11. MORAN, L. J., MORAN, FRANCES A., & BLAKE, R. R. An investigation of the vocabulary performance of schizophrenics: II. Conceptual level of definitions. *J. genet. Psychol.*, 1952, **80**, 107–132.
12. MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67**, No. 20 (Whole No. 370).
13. RAPAPORT, D. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.
14. SULLIVAN, H. S. The language of schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 4–16.
15. WECHSLER, D. *The measurement of adult intelligence*. Baltimore: Williams & Wilkins, 1944.

Received January 17, 1955.

EXPECTATIONS, FULFILLMENT, AND MORALE[1]

AARON J. SPECTOR

Officer Education Research Laboratory, Maxwell Air Force Base, Alabama

THERE exists today a large body of knowledge pertaining to the bases for the attractiveness of certain goal activities. It is commonly accepted, for example, that deprivation of a substance biologically required results in that substance's becoming a goal object, viz., attractive and satisfying to consume. In illustration, we frequently refer to the chalk craving of calcium-deprived children. Similarly, the peculiar diet of some primitive tribes and the effects of hunger and thirst have also been cited as evidence of the role that biological deprivation plays in affecting the attractiveness of goal objects.

A number of investigators have recently explored the relationships between goal attractiveness and such variables as: physical proximity (10, 11), probability of attainment (5, 6, 17); valence of members associated with the goal (8), personal ownership and ownership by others (12). Since the results of some of these studies bear on the interpretation of data presented below, they will be discussed later.

The research of learning theorists on the concepts of goal gradients, acquired drives, reinforcement, and stimulus generalization have contributed significantly to our knowledge in this area. Although most of their work has been done on lower-order animals, there is a definite trend toward the application of learning theories to complex human situations (2, 21, 22, 24). However, the traditional theories are frequently inadequate for the interpretation and prediction of real-life goal activities, as evidenced in numerous morale studies.

For instance, in one study conducted by the Army, it was discovered that the men who were most dissatisfied with the promotion system were in branches which had the highest rate of promotion, viz., the Air Force, while men in the Military Police, which had the lowest rate of promotion, were the most satisfied with the promotion system. To explain this anomalous finding the Army researchers resorted to their newly formulated *relative deprivation hypothesis*,[2] which is based on the principle that expectations are founded in reference groups, and satisfactions are to some degree functions of these expectations.

According to Merton and Kitt (20) the hypothesis was used on nine different occasions in the researchers' monumental report (26), and in every case it was invoked after the data were in to explain some enigma. The evidence mustered by the researchers, on an ex post facto basis, in support of their hypothesis was rather convincing, albeit too lengthy to review here.

Examination of the relative deprivation principle reveals that it has two aspects: (*a*) the bases of expectations,[3] and (*b*) the satisfactions following fulfillment of these expectations. The research herein reported has been addressed to the latter aspect. It represents an attempt to test experimentally an aspect of the relative deprivation hypothesis in a context simulating one studied by the Army researchers. The independent variables necessarily were dictated by the definition of the problem, and were as follows: (*a*) Expectations (perceived probability of promotion in a pseudo-military hierarchy), and (*b*) Fulfillment (promotion or nonpromotion). The dependent variable was morale.

The effects on morale of expectations of promotion and the degree of fulfillment of these expectations, have never been experimentally investigated despite their relevance to our society. These factors operate in the military and in industry; in fact, they operate in all hierarchical social structures and conceivably may affect the morale of the organization's members.

[1] The author was on the faculty of the University of Massachusetts when this article was written. The article is based on part of a dissertation submitted to the Department of Psychology, Boston University, in partial fulfillment of the requirements for the Ph.D. The author wishes to express his indebtedness to N. Maccoby and B. Hymovitch for direction and aid throughout the course of this research. A summary of this paper was read at the 1953 Convention of the American Psychological Association.

[2] In psychological literature the term deprivation usually refers to the absence of substances needed for life, but this is not implied in the hypothesis. We will substitute the term *frustration* for the term *deprivation* in our own version of this hypothesis.

[3] The sociological aspects of expectations are explored extensively by Merton and Kitt (20).

The concept of expectations is continually assuming greater importance in psychological literature although major difficulties still exist in isolating expectations experimentally—a fact made clear in several studies (4, 19, 23)—and measurement poses still other problems (9, 17, 24).

In the theories of Tolman and Krech, expectations are considered a major theoretical construct. MacCorquodale and Meehl (18) are in the process of developing an expectancy learning theory. Similarly, Rotter (24) and his students are working on a "social learning theory which combines both the constructs of expectancy and reinforcement." Woodworth now uses the concept of expectations and contends that it is the expectancy which is reinforced in conditioning studies, while Hull explains anticipatory behaviors of rats by his concept of fractional antedating goal responses. The role of expectations and success or failure experiences in level of aspiration studies has been the subject of numerous researches over the past two decades.

The concept of reinforcement is firmly entrenched in contemporary theories despite differences of opinion over its drive-reduction quality, as well as the question of what is reinforced. The latter issue is again raised by the present study, if we are justified in interchanging the terms fulfillment and reinforcement.

HYPOTHESES

The present research is addressed to only one aspect of the Army researchers' hypothesis, namely, the manner in which expectations and fulfillment relate to satisfactions. To this end, a *frustration hypothesis* will be defined as follows:

Hypothesis I. On failing to achieve an attractive goal, an individual's morale will be higher if the probability of achieving that goal had been perceived to be low than if it had been perceived to be high.

The first hypothesis could easily have been derived from the fields of learning or personality theory since it is simply a statement of the effects of frustration through nonreward. The next hypothesis, which might be called a *gratification hypothesis*, is not as directly derivable from any systematic theory although it is popularly accepted as a common-sense truism.

Hypothesis II. On achieving an attractive goal, an individual's morale will be higher if the probability of achieving that goal had been perceived to be low than if it had been perceived to be high.

These two hypotheses suggest that regardless of the individual's success or failure in achieving a goal, his morale will be partly determined by his perceived probability of successful achievement. This *expectation hypothesis* is a general hypothesis of which the *frustration* and *gratification hypotheses* are subhypotheses.

Hypothesis III. On succeeding, or failing, to achieve an attractive goal, an individual's morale will be higher if the probability of achieving that goal had been perceived to be low than if it had been perceived to be high.

From common sense and a multitude of empirical observations one can deduce the hypothesis that achieving an attractive goal would result in higher morale than failing to achieve that goal. This *fulfillment hypothesis* might be stated as:

Hypothesis IV. For all given perceived probabilities of achieving an attractive goal, an individual's morale will be higher if he achieves that goal than if he fails to achieve it.

On the basis of Hypotheses III and IV we can make six predictions about the rank ordering of morale scores of different groups when the following assumptions are met: the goal in question is attractive to all the individuals; these individuals are equally desirous of achieving the goal; and certain individuals perceive the probability of achieving the goal as low, while others perceive the probability as high.

For the sake of brevity, (L) and (H) will signify *S*s who perceived the probability of promotion as Low and High respectively; (P) and (NP) signify *S*s who were Promoted or Not Promoted. The specific predictions about magnitude of morale scores of the four treatments are:

Treatment:	L:P	L:P	L:P	H:P	H:P	L:NP
Higher morale than:	H:P	L:NP	H:NP	L:NP	H:NP	H:NP

PROCEDURES

Subjects. Thirty-six groups, of four men each, were randomly assigned to the four treatments in a 2 × 2 design.[4] In six groups a stooge took the place of one

[4] All *S*s were enrolled in the sophomore general psychology course at the University of Massachusetts.

man who failed to appear, which reduced the *N* to 138.

Instructions. The instructions served to orient *S*s toward the problem of teamwork in military intelligence. They were to operate as a team in which each man was to decode a different part of the group's message. The *S*s were seated in a rectangle formed by four desks on which small dressing-room shields were placed to keep each man hidden from the others. Two slots were cut into the bottom of each shield which allowed written messages to be passed out and in. The experimenter stood in the middle of the rectangle and delivered the notes by sliding them into the slots.

Tasks. Each team worked on two messages, the second of which was slightly more complex than the first one. The messages were written in *codes* and were soluble by direct substitutions from the *code sheet.* The letters of each message were divided into eight columns, six rows long; each man was to decode two columns. When he completed decoding four letters, he was to send notes to the other men telling them what his letters stood for.[6]

Probability of promotion. The *S*s were divided into two groups, as follows: Group H was told that three out of the four men would be promoted from corporal to top sergeant after the first message was completed; Group L was told that one out of the four men would be promoted. The experimenter emphasized the status aspects of the promotion in order to ensure that promotions were attractive to all people. It was thus assumed that Group H had a high expectation of promotion, Group L a low expectation.

Fulfillment. The *S*s were further divided on the basis of their promotion or nonpromotion in the experiment. All *S*s in the P groups were promoted; all *S*s in the NP groups failed to achieve promotion. To minimize artificiality in the experiment the *S*s were encouraged to familiarize themselves with each other via autobiographical inventories which were passed around. They were also allowed to cast ballots for promotions, but were not aware that the experimenter had entirely disregarded their votes.

A measure of the success of these instructions, and of *S*s' motivation to be promoted, was obtained by administering a questionnaire after the first message was completed. At the same time the *S*s checked the names of the man/men they wanted promoted.

Morale. A six-item morale scale was administered after completion of the second message. Each item was checked on a seven-point continuum which ran from "highly satisfied" to "highly dissatisfied." Lower scores signified a more favorable attitude toward that item.

The six questions measured the following aspects of morale: (I) the group's attractiveness to its members, (II) satisfaction with the task of decoding the first message, (III) desire to remain a member of the group, (IV) satisfaction with the promotion system, (V) evaluation of the group's productivity, (VI) satisfaction with task of decoding the second message.[6]

[5] This technique of sending notes in order to control communications has been used by many others, e.g., Kelley (15), Thibaut and Coules (27), and Grossack (7).

[6] These dimensions of morale were adapted from the Survey Research Center studies (13, 14).

RESULTS

The instructions were designed to make promotions attractive to all *S*s regardless of the probability of a promotion. An inquiry on the first questionnaire (prior to introduction of the Fulfillment variable) measured the attractiveness of the goal, promotion, to *S*s in the four treatments. Analysis of variance of this question yielded an *F* ratio of .07. Evidently, motivation was not influenced by the different probabilities of promotion.[7] However, this might have been an indirect result of instructions too weak to influence a person's expectations of getting a promotion. The appropriate question then is: Did a significantly higher percentage of the H groups expect a promotion to sergeant than the L groups?[8] The chi-square difference was found to be beyond the .01 level (Table 1); the H groups expected promotions, whereas the L groups did not.

Pure cases. The instructions were intended to induce varied perceptions about promotion opportunities between treatments, but unfortunately the instructions were not completely successful, as shown in Table 1. We shall call those people who did react as intended, the *pure cases.*

It is conceivable that people in the H groups who perceived the probability of their getting a promotion as low (contrary to the instructions), might react in a manner similar to the people actually in the Low Probability groups. The available evidence supports this conjecture. When mean scores of all people in the H and L groups were compared, the differences between treatments were smaller than when only *pure cases* in the treatments were compared. Apparently the results could be confounded by subjects whose expectations were not veridical with those intended for them by the experimenter. In view of the facts that (*a*) the hypotheses relate only to *S*s whose expectations were of a certain specified nature, and (*b*) differences between treatments were altered when data were included of *S*s whose expectations were not congruent with the expectations

[7] This finding supports Rotter's (24) presumption of independence of motivation and expectations.

[8] Although this method of measuring subjective expectations might have been complicated by motivational variables, as others have found (4, 19, 23, 24), it was relatively less contaminated than other methods tried in pretests. A comparison of various techniques for measuring expectations has been reported by Rotter (24).

TABLE 1

CHI-SQUARE DIFFERENCES BETWEEN H AND L TREATMENTS ON THEIR EXPECTATIONS OF PROMOTION*

N = 138		Treatment H	Treatment L
Expected	Sgt.	41	12
	Cpl.	25	60

* "Which position do you expect to have on the next message? Corporal_____, Sergeant_____."

Note.—Chi square = 36; df = 1; $p < .01$.

TABLE 2

ANALYSIS OF VARIANCE OF INDEX OF GENERAL MORALE, BASED ON SIX QUESTIONS

Source of Variation	df	Mean Square	F	p
Perceived probability	1	76.19	4.07	$<.05$
Fulfillment	1	157.53	8.42	$<.01$
Interaction	1	25.80	1.38	
Within groups	97	18.69		

of others in their treatment, we are justified in testing the hypotheses with the data of only *pure cases*. Therefore, only the data based on *pure cases* will be reported in this paper.

Tests of hypotheses. In order to test the major hypotheses, an index of general morale was constructed. The scores to each of the six questions for each individual in a particular treatment were summed and a treatment mean obtained. The rank order of the mean morale scores was exactly as predicted by the hypotheses: (*a*) L:P, (*b*) H:P, (*c*) L:NP, (*d*) H:NP. If chance alone were operating, this particular order of means would be expected once in 24 times. Furthermore, morale was higher among the *Ss* in Group L than in Group H, and among the *Ss* in Group P than NP, as Hypotheses III and IV predict.

The scores of this index were submitted to analysis of variance to test the null hypothesis of no difference between means. The F ratio that was produced was significant beyond the .01 level. These scores were then submitted to further analysis based upon a partitioning of sums of squares between groups.[9] In this analysis Fulfillment was found to produce an F significant beyond the .01 level of confidence, while Perceived Probability was significant at the .05 level (Table 2).

Six predictions have been made about the relative magnitude of the groups' satisfaction scores on each question. The rankings on Questions I and V were perfectly predicted, while for each of the remaining questions, five of the six predictions were correct (Table 3).

The null hypothesis of no difference between group means on each of the questions was tested by submitting the scores to each question, for all treatments simultaneously, to analysis of variance. Three of the six F's thus obtained were significant. These questions were submitted to further analysis based upon a partitioning of sums of squares between groups as shown in Tables 4, 5, and 6.

Fulfillment was a significant determinant of attitudes toward the group (Question I) and toward the second task (Question VI). Attitudes toward the group were also influenced by the Perceived Probability of promotion. Neither of the variables yielded a significant F ratio on Question V, evaluation of the group's productivity, when only *pure cases* were analyzed. However, when data on this question

TABLE 3

MEANS AND STANDARD DEVIATIONS OF ALL TREATMENTS FOR EACH QUESTION

Question		L:P	H:P	L:NP	H:NP
I	Mean	2.37	2.71	2.83	3.30
	σ	.98	.89	1.01	.95
II	Mean	2.24	2.33	2.77	2.75
	σ	1.00	1.17	1.56	.99
III	Mean	2.51	2.52	2.52	3.10
	σ	1.11	.91	1.11	1.41
IV	Mean	2.55	3.14	2.87	3.60
	σ	1.81	2.00	1.59	2.00
V	Mean	3.13	3.23	3.29	3.90
	σ	1.06	.72	1.14	.30
VI	Mean	2.41	2.14	2.71	3.25
	σ	1.30	1.09	.99	.94
	Mean	2.53	2.67	2.83	3.31

TABLE 4

ANALYSIS OF VARIANCE OF QUESTION I—ATTRACTION TO MEMBERS OF THE GROUP

Source of Variation	df	Mean Square	F	p
Perceived probability	1	3.81	3.96	$<.05$
Fulfillment	1	6.53	6.80	$<.01$
Interaction	1	.66	.68	
Within groups	97	.96		

[9] The sums of squares may be "partitioned into as many component parts as there are degrees of freedom associated with it" (3) in a 2 × 2 factorial design such as this.

TABLE 5

ANALYSIS OF VARIANCE OF QUESTION V, USING SCORES OF THE TOTAL POPULATION

Source of Variation	*df*	Mean Square	*F*	*p*
Perceived probability	1	1.39	1.68	
Fulfillment	1	5.16	6.26	<.01
Interaction	1	.11	.13	
Within groups	134	.82		

TABLE 6

ANALYSIS OF VARIANCE OF QUESTION VI—SATISFACTION WITH TASK OF DECODING SECOND MESSAGE

Source of Variation	*df*	Mean Square	*F*	*p*
Perceived probability	1	.38	.30	
Fulfillment	1	9.81	7.84	<.01
Interaction	1	3.91	3.12	
Within groups	97	1.25		

from the total population were analyzed, Fulfillment was found to be significant beyond the .01 level (Table 5).

Fulfillment also yielded a significant *F* ratio (.05) on analysis of an *Index of Attitudes Toward the Group*, composed of Questions I and III, and an *F* ratio at the .01 level on an *Index of Task Satisfaction*, composed of Questions II and VI.

DISCUSSION

This research has been concerned with an experimental attempt to verify an aspect of the relative deprivation and several related hypotheses. In essence, these hypotheses stated that either low probability of occurrence of an attractive event, or the actual occurrence of that event, would result in greater satisfaction than high probability, or the nonoccurrence of the event.

Perceived probability. Among those who were actually Promoted (P), the *S*s who perceived the probability of promotion as low (L) responded more favorably than did *S*s who perceived the probability as high (H). This was true on five of the six items of the questionnaire, which supports our *gratification hypothesis*.

Similarly, the *frustration hypothesis* was supported by data on those who were Not Promoted (NP). Among the NP groups, the responses of L groups were more favorable than those of H groups on five of the six questions.

A significant *F* ratio for Perceived Probability was obtained on the index of general morale. Reference to Table 3 reveals that the means were all in the order predicted. These facts, together with the verification of the *frustration* and *gratification hypotheses*, militate toward acceptance of the more general *expectations hypothesis*. Since these results were obtained in a laboratory experiment, where the expectations were artificially induced, it is quite reasonable to believe that expectations in everyday life play an even greater role in determining satisfactions.

Somewhat similar hypotheses have been tested in other studies, each time with conflicting results. On the one hand, Gebhard's (6) results are all consistent with those reported here, while in the case of Filer (5), only the gratification hypothesis was supported, and in the case of Klass (16), only the frustration hypothesis was supported. It has been suggested that the use of a "weak" fulfillment variable precluded the possibility of finding support for the gratification hypothesis in Klass' study (25), whereas Filer attributes his own rejection of those hypotheses involving expectations to the fact that his instructions did not create differential expectations in his subjects.

Fulfillment. Even more strongly indicated as a factor in morale is the variable of Fulfillment, viz., promotion or nonpromotion. In every analysis made, the people who were promoted responded more favorably than those who were not, which is what the *fulfillment hypothesis* would lead one to expect. These results, as well as those of the Army's survey, suggest that personnel managers might be wise to underplay, rather than overplay, the opportunities for advancement in their organizations. An investigation of this problem under industrial conditions might be fruitful.

The data cited above make it quite apparent that a person's satisfaction with his state of affairs is partly a function of how certain he felt this state would come about. Morale does not appear to be adversely affected by failure to achieve an attractive goal if the person's expectations of achieving the goal were low. Low expectations seem to act as a buffer to morale when an attractive event does not come to pass. Its effect is equally salubrious when the event does occur.

If we translate these results into learning theory terms, we may address ourselves to at least one major point of contention between the positions taken by Tolman, Hull, and Guthrie. We recall that the highest morale scores were associated with the low expectation groups who were ultimately promoted. The question arises, then: What was reinforced by the promotion? Was it the low expectations? or the responses on the task? or the voting for promotions?

Following Guthrie, we would expect the responses immediately preceding the reward, i.e., voting for promotions, to be more attractive to the Promotion groups than the Nonpromotion groups. The results contraindicate such an hypothesis, since the high expectations-promotion group was less satisfied with the promotion system than was the low expectations-nonpromotion group. The results are more disposed toward a Hullian approach in that the Promotion groups indicated greater satisfaction with the goal responses, i.e., the coding tasks, than did the Nonpromotion groups. Nevertheless, it is difficult to see how Hullian or any other learning theory would cope with some of the data, such as the satisfactions with the promotion system. The difficulty probably lies in our inadequate understanding of the influence of expectations on man's evaluation of himself, and on his perception of goal barriers, in success and failure situations. It is to be hoped that the newer expectation theories (18, 24, 27) will be more adequate for the complexities of social psychological research.

REFERENCES

1. CHILD, I. L., & ADELSHEIM, E. The motivational value of barriers for young children. *J. genet. Psychol.*, 1944, **65**, 97–111.
2. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York, McGraw-Hill, 1950.
3. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
4. FESTINGER, L. Wish, expectation, and group standards as factors influencing level of aspiration. *J. abnorm. soc. Psychol.*, 1942, **37**, 184–200.
5. FILER, R. J. Frustration, satisfaction, and other factors affecting the attractiveness of goal objects. *J. abnorm. soc. Psychol.*, 1952, **47**, 203–212.
6. GEBHARD, MILDRED E. Changes in the attractiveness of activities: The effect of expectation preceding performance. *J. exp. Psychol.*, 1949, **39**, 404–413.
7. GROSSACK, M. M. Controlling interaction in small group research, *J. Psychol.*, 1953, **35**, second half, 241–244.
8. HOROWITZ, M. N., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1950, **4**, 57–77.
9. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
10. IRWIN, F. W., ARMITT, F. M., & SIMON, C. W. Studies in object-preferences. I: The effect of temporal proximity. *J. exp. Psychol.*, 1943, **33**, 64–72.
11. IRWIN, F. W., ORCHINIK, C. W., & WEISS, J. Studies in object-preferences: The effect of temporal proximity upon adults' preferences. *Amer. J. Psychol.*, 1946, **59**, 458–462.
12. IRWIN, F. W., & GEBHARD, MILDRED E. Studies in object preferences: The effect of ownership and other social influences. *Amer. J. Psychol.*, 1946, **59**, 633–651.
13. KATZ, D., MACCOBY, N., & MORSE, N. C. *Productivity, supervision and morale in an office situation*, Ann Arbor: Univer. of Michigan, 1950.
14. KATZ, D., MACCOBY, N., GURIN G., & FLOOR, L. G. *Productivity, supervision and morale among railroad workers*, Ann Arbor: Univer. of Michigan, 1951.
15. KELLEY, H. H. Communication in experimentally created hierarchies, *Hum. Relat.*, 1951, **4**, 39–56.
16. KLASS, B. The role of expectations in determining intrinsic job satisfaction. Unpublished doctoral dissertation, Boston Univer., 1952.
17. LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
18. MACCORQUODALE, K., & MEEHL, P. E. Preliminary suggestions as to formalization of expectancy theory. *Psychol. Rev.*, 1953, **60**, 55–63.
19. MARKS, ROSE W. The effect of probability, desirability, and "privilege" on the stated expectations of children. *J. Pers.*, 1951, **19**, 332–351.
20. MERTON, R. K., & LAZARSFELD, P. F. (Eds.) *Continuities in social research*. Glencoe, Illinois: The Free Press, 1950.
21. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
22. MOWRER, O. H. *Psychotherapy: Theory and research*. New York: Ronald, 1953.
23. PRESTON, M. G., & BAYTON, J. A. Differential effect of a social variable upon three levels of aspiration. *J. exp. Psychol.*, 1944, **29**, 351–369.
24. ROTTER, J. B., FITZGERALD, B. J., & JOYCE, J. N. A comparison of some objective measures of expectancy. *J. abnorm. soc. Psychol.*, 1954, **49**, 111–114.
25. SPECTOR, A. J. Expectations, outcomes and satisfactions. Unpublished doctoral dissertation, Boston Univer., 1953.
26. STOUFFER, S. A., SUCHMAN, E. A., DE VINNEY, L. C., STARR, S., & WILLIAMS, R. M., JR. *The American soldier; adjustment during army life*, Vol. 1, Princeton: Princeton Univer. Press, 1949.
27. THIBAUT, J. W., & COULES, J. The role of communications in the reduction of interpersonal hostility, *J. abnorm. soc. Psychol.*, 1952, **47**, 770–778.
28. WRIGHT, H. F. The influence of barriers on strength of motivation. *Contr. Psychol. Theory*, 1937, **1**, 1–143.

Received January 17, 1955.

THE PERCEPTION OF SOCIAL ATTITUDES[1]

SAMUEL J. MESSICK[2]
University of Illinois

THE attitudes of others may be perceived as organized in different ways by different groups of people. Consider an attitude area composed of attitude toward war, attitude toward capital punishment, and attitude toward the treatment of criminals (correction vs. punishment). Certain people may perceive others as varying from an extreme humanitarian to a correspondingly extreme nonhumanitarian viewpoint. In such a situation some particular person might be seen as being antiwar, anticapital punishment, and procriminal correction in the sense that one attitude directly implies the other and that no alternatives are possible (see Fig. 1). Such a person might be called a "humanitarian." In the same type of schema, some other person, who would be called inhumane, might be seen as being prowar, procapital punishment, and anticriminal correction. This would be a one-dimensional system, since the attitudes are not perceived separately but as one composite "humane-inhumane" attitude. People with such a point of view could not conceive of a person who was both antiwar and procapital punishment.

Another group of people may perceive the attitudes of others as being structured in a different way (see Fig. 2). In this type of structure capital punishment of extreme criminals is equated with punishment of criminals in general as a deterrent to crime, but war is seen as something independent. This dimensional configuration provides a "frame of reference" in which Type B finds it possible to conceive of people in each of the four quadrants. In this perceived structuring, two pacifists could be conceived of as having widely different views toward capital punishment. Although they are both viewed as being against war, one could be seen as in favor of capital punishment as a protection to society and a deterrent to crime and the other as against it because "it is just as uncivilized as war." Type B would also feel it possible for two people to be violently anticapital punishment and yet have different views toward war, because war is considered to be essentially independent of attitudes toward criminals. It would be difficult under these circumstances, however, for Type B to conceive of a person who was in favor of punishing minor criminals and at the same time against capital punishment for extreme criminals.

In Type C war and capital punishment are equated as two means of taking human life, but criminal correction is seen as something independent. Thus, Type C might conceive of two religious conscientious objectors as being against both war and capital punishment because killing is involved, and yet one may be considered procorrection for humanitarian reasons and the other propunishment because evil-doers must pay.

For Type D these three attitudes are viewed as being independent of each other, and it is possible for this group to conceive of people in each of the octants. Of course, the dimensions in any of the above types may be related in various ways—the lines drawn at angles instead of perpendicularly.

It must not be construed from the above discussion that a given group of people perceive the attitudes of others in only one way. The above types may be considered to be "points of view," and it is certainly possible for an individual to hold several different points of view with respect to the attitudes of others, depending upon the characteristics of the "other" people considered. Some individuals may be able to conceive of a different attitude structure for every group of people specified, because they perceive different types of people in different ways. However, it is also possible that these individuals perceive one particular attitude structure more often than another,

[1] This study was supported in part by the Office of Naval Research contract N6onr-270-20 with Princeton University and also in part by funds from the National Science Foundation. The opinions and conclusions expressed are, of course, those of the author and do not represent the views of the Office of Naval Research or of the National Science Foundation. The initial portion of this study was carried out when the author was an Educational Testing Service Psychometric Fellow at Princeton University.

[2] Postdoctoral Fellow for Research in Personality, Ford Foundation Grant to the University of Illinois.

and that in the absence of specific knowledge or a detailed characterization of the other person, they may ascribe to his attitudes this modal structure. It is this modal, or dominant, way in which individuals generally structure the attitudes of others with which the present study is concerned.

In psychophysics multidimensional scaling methods (12, 13) have been developed which yield dimensional configurations for psychophysical judgments. These methods may be promising for the analysis of attitudes. The fundamental concept involved in these techniques is psychological distance, a construct that has been widely used by many psychological theorists. Psychological distance in the attitude realm can be thought of in terms of the degree of agreement or disagreement shown between two statements. If a person who strongly agrees with one statement would also be likely to agree strongly with another, then these two statements can be considered to be psychologically close together. If a person who strongly agrees with one statement would not be very likely to agree with the other, then those two statements can be considered to be psychologically far apart. If the psychological distances among attitude statements can be analyzed in a Cartesian space, it will be possible to obtain a configuration of the way in which an individual perceives attitudes as being structured in a given domain.

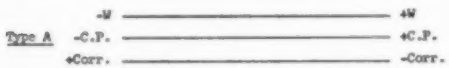


FIG. 1. CONCEPTUALIZATION OF THE ATTITUDES TOWARD WAR, TOWARD CAPITAL PUNISHMENT, AND TOWARD TREATMENT OF CRIMINALS AS A UNIDIMENSIONAL SYSTEM

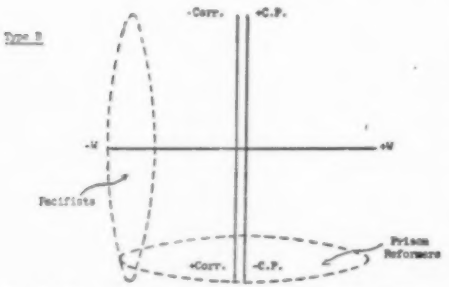


FIG. 2. CONCEPTUALIZATION OF THE ATTITUDE TOWARD CAPITAL PUNISHMENT COMBINED WITH THE ATTITUDES TOWARD CORRECTION OF CRIMINALS WHILE THE ATTITUDE TOWARD WAR IS SEEN AS AN INDEPENDENT DIMENSION

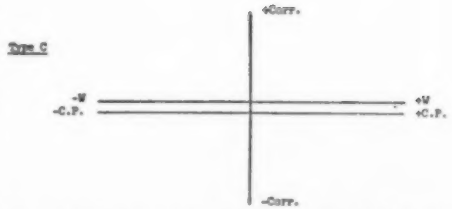


FIG. 3. CONCEPTUALIZATION OF ATTITUDES TOWARD WAR AND CAPITAL PUNISHMENT COMBINED AS TWO MEANS OF TAKING HUMAN LIFE WHILE THE ATTITUDE TOWARD CORRECTION OF CRIMINALS IS SEEN AS INDEPENDENT

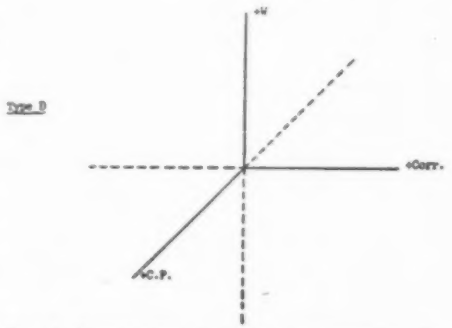


FIG. 4. CONCEPTUALIZATION OF THE ATTITUDES TOWARD WAR, TOWARD CAPITAL PUNISHMENT, AND TOWARD THE CORRECTION OF CRIMINALS AS THREE COMPLETELY INDEPENDENT DIMENSIONS

THE PROBLEM AND AN APPROACH TO ITS SOLUTION

As was seen in the introduction, it is possible on a priori grounds to structure the attitudes of others in terms of several different, yet reasonable, dimensional arrangements. The actual attitude arrangements perceived by individuals, however, has seldom been investigated, mainly because of a lack of appropriate methods. Knowledge of the ways in which individuals perceive attitudes as being structured in others is important psychologically, since such perceptions affect an individual's understanding of others, his relationships and adjustments with others, and his actions toward others. It might be possible to obtain some information about these perceived atti-

tude structures by utilizing multidimensional scaling techniques.

The purpose of this study, then, is first of all to see whether a set of perceived attitude relationships can be adequately represented in dimensional terms and, secondly, to see if two groups, which probably differ with respect to these attitudes, perceive them as being structured in different ways. The adequacy of the dimensional representation can be evaluated in terms of the requirements of the multidimensional psychophysics model (see 7, 13, 18). This model provides for a Euclidean dimensionalization of the data, and if the data do not have Euclidean properties, one of the requirements of the model will not be met, i.e., the matrix of distance scalar products (see 7, 18), which must be positive semi-definite for Euclidean distances, will be found to have large negative roots. Therefore, if large negative roots are found, the suitability of the model is in question, and dimensions should not even be extracted. Unfortunately, however, statistical tests for the size of latent roots are not available, so arbitrary acceptance regions must be set (7, 13). Under such conditions, it might sometimes be possible for a matrix to be dimensionalized when it is actually non-Euclidean. Even in this case, however, there are indications of the unsuitability of the model, for a forced Euclidean description of non-Euclidean distances would lead to an increase in dimensionality. Thus, whenever the number of dimensions extracted approaches the number of stimuli, not only is the fruitfulness of the approach in that area in doubt, but the suitability of the model is also in question.

Multidimensional scaling methods have been utilized by Richardson (method of triadic combinations [10]) and Torgerson (method of complete triads [13]) which could be applied for this purpose. These methods are essentially extensions of the method of paired comparisons to the multidimensional case. The method of complete triads could be applied to attitude statements as it stands, but it has the undesirable property that for n statements $n(n-1)(n-2)/2$ judgments would be required from each S. The task becomes prohibitive with more than 10 statements. It would be more reasonable to apply a multidimensional extension of the method of successive intervals (3, 11), which would require only $n(n-1)/2$ judgments. The procedure for such a method would require S to arrange $n(n-1)/2$ pairs of stimuli on a distance continuum according to the degree of similarity of the members of each pair. This question of similarity or relatedness can be approached in the attitude realm by asking for a judgment of the agreement or disagreement shown between two statements. The judgment involved can be set up in terms of the attitudes of some *other* person in the following way: S is asked to imagine the type of person who would strongly agree with Statement A and then to decide how this same person would also feel about Statement B. In this way, S is asked to estimate the attitude relationships of others, and the multidimensional analysis of these data will provide a dimensional configuration of these perceived attitude relationships. Minimal information about the "other" person is provided in this procedure, i.e., he is characterized only as being in complete agreement with a single statement. Perhaps, in this highly unstructured situation, this agreement will act as a cue to the complex characterization which S most often ascribes to other people in general, and S will respond in terms of this modal, or dominant, characterization.

Since there is a persistent correlation between a person's own attitudes and his estimates of the attitudes of others (5, 14, 16), the procedure suggested here for investigating the structure of perceived attitudes bears certain similarities to projective instruments. However, such correlations characteristically range from .4 to .6. Such a relationship seems to indicate the possibility of assessing a person's own attitudes by giving him the seemingly objective task of estimating the opinions of others, but this is not the intent of the present experiment. The suggested procedure as it stands does not necessarily yield the structure of S's own attitudes or the actual structure of attitudes in others, but it does yield information about the ways in which people *perceive* attitudes "out there" in others. Two groups were used in the present experiment, which were thought to be diverse with respect to the attitudes involved. Any differences in the attitude configurations obtained from these groups might possibly be traced not only to differences in perception but also to differences in the Ss' actual attitudes which have been pro-

jected into the situation. Since the primary concern of the present study was with the perception of attitudes, the directions and the task were so formulated as to minimize the effects of projection. The success of this intention and the projective aspects of the task as formulated will be considered in a later section.

METHOD

In the present study a multidimensional method of successive intervals (1, 6) was applied to a set of attitude statements in an attempt to discover the number and the nature of the relevant dimensions perceived. The attitude statements were chosen from three of Thurstone's scales—attitude toward war, attitude toward capital punishment, and attitude toward the treatment of criminals. The procedure was applied using two diverse groups in an attempt to discover any differences which may exist in the structuring of these perceived relationships. The groups chosen for this purpose were a theological seminary class and an Air Force officer candidate class.

The Stimulus Statements and Their Experimental Presentation

Seven statements were selected from each of the three Thurstone scales on attitude toward war (9), attitude toward the treatment of criminals (17), and attitude toward capital punishment (8). These statements, which were selected in such a way as to be fairly evenly distributed over the favorable-unfavorable continuum for each of the three attitudes, are listed below.

1. On the whole, wars do the world some good.
2. Peace and war are both essential to progress.
3. War is sometimes necessary because right is more important than peace.
4. War has some benefits, but it's a big price to pay for them.
5. Although war is terrible, it has some value.
6. War is hardly necessary in the modern world.
7. There is no conceivable justification for war.
8. All criminals are victims of circumstances and deserve to be helped.
9. One way to deter men from crime is to make them suffer.
10. Correction is more effective than punishment in preventing crime.
11. Harsh imprisonment merely embitters a criminal.
12. Criminals must be punished.
13. Petty offenders can be reformed without punishment.
14. Failure to punish the criminal encourages crime.
15. The death penalty should be given for major crimes.
16. Capital punishment is good because it deters others from crime.
17. I think capital punishment is necessary but I wish it were not.
18. Until we find a more civilized way to prevent crime we must have capital punishment.
19. No thinking individual can believe in capital punishment as a method of preventing crime.
20. Criminals are pathological people who should be corrected, not executed.
21. The death of a comrade in prison embitters all the inmates against the state.

In this procedure, it is not at all necessary to use previously scaled attitude statements. Any other set of statements, scaled or otherwise, which was thought to be of possible relevance to the domain in question could have been used. The selection of previously used items for the present investigation was for the purpose of gaining some incidental information about the claimed unidimensionality of the Thurstone scales.

The statements were arranged in booklet form in order to make the task more amenable to group presentation. One statement was printed at the top of each page, followed by a list of statements at the left. To the right of each of these statements appeared a series of 11 boxes, with the first box labeled "Strongly Disagree", the sixth box labeled "Neutral", and the eleventh box labeled "Strongly Agree". The *S* was asked to imagine the type of person who would strongly agree with the statement at the top of the page and then to decide how this same person would also feel about each of the statements on the left. The *S* was then to indicate the extent of this person's agreement or disagreement with each statement by placing a cross in one of the boxes to the right. In constructing the booklet the order of appearance of the statements on each page and the order of the pages in the booklet was determined by random number techniques.

The Judgments Required

The multidimensional method of successive intervals requires that the subject arrange all possible pairs $[n(n-1)/2 = 210]$ of the 21 statements on a distance continuum according to the degree of relatedness or agreement shown between the members of each pair. In order to obtain judgments about perceived attitudes, the question of relatedness was set up in terms of the attitudes of some other person as described above. In terms of the multidimensional model this task can be considered as asking *S* to estimate the distance between Statements A and B, while located at A and looking at B. The reverse judgment for each pair of statements was also obtained in the present experiment, i.e., *S* was also asked to imagine the type of person who would strongly agree with Statement B and then to decide how this same person would also feel about Statement A. This task may be considered as judging the distance from Statement B to Statement A, while located at B and looking at A. Thus, the procedure requires two judgments for each of the 210 pairs of statements, making a total of 420 judgments required from each subject.

The two judgments of the same distance taken from opposite directions were obtained in the present experiment in order to see whether the data met one of the basic requirements of the multidimensional scaling model. The multidimensional attitude scaling model is an interpoint-distance model, i.e., attitude statements and people are represented in the same space by points. The extent of the individual's agreement or disagreement with the statements in the space is represented by the distance between the person and the statement. The person agrees with items close to him in any direc-

tion and disagrees with items far away. One property of this model, then, is that the probability of an affirmative response to any statement decreases as the distance between the statement and the person increases, without regard to direction. Statements selected by Thurstone scaling methods such as equal-appearing intervals, paired comparisons, or successive intervals usually have this property. Such statements are commonly referred to as "point" statements, and the multidimensional scaling model can be properly applied only to such items. The requirement that the probability of an affirmative response should vary inversely with distance independent of direction is another way of saying that the distance between two points in Euclidean space should be the same in both directions.

In the present experiment, then, an attempt was made to see whether judgments of the same distance taken from opposite directions would come out similar. The two judgments elicited for each pair of statements were considered as independent estimates, and scale values were obtained for both the distances AB and BA for each pair. If the multidimensional scaling requirement that the distance between two points be the same in both directions is met by the data, the differences between corresponding AB and BA distances should be small. Systematic groupings of large differences would probably indicate the type of asymmetric relationships that might reasonably be expected to occur with cumulative items (4) as opposed to the symmetric relationships expected from point items. If a few large differences occur, then, the statements involved in these differences should be deleted from the study on the grounds of a failure to meet certain requirements of the model. However, great care should be exercised in evaluating these differences, since a few large discrepancies might occur by chance alone. Also, large differences would be expected for statement pairs at either extreme of the similarity scale, since such scale values are based on only a few cases and, hence, are very unstable.

The Subjects

The two diverse groups selected for this experiment were a seminary class and an Air Force officer candidate class. The seminary group consisted of 40 third-year male students at the Princeton Theological Seminary in Princeton, New Jersey. A regularly scheduled class hour had been made available for the purposes of this experiment, and the enlisting of *Ss* was put on a purely voluntary basis. Forty volunteers were obtained from the class, and the rest of the students were permitted to leave the room.

The Air Force group consisted of 82 male officer candidates who had just finished the first half of their O.C.S. training at Lackland Air Force Base, San Antonio, Texas.

No direct measure of the actual attitudes of these two groups was obtained in the present investigation for several reasons. Instead, attitude differences were inferred from the widely different backgrounds of the *Ss*. It seemed reasonable that factors which had led to such divergent career choices might also be reflected in divergent attitudes, especially in an attitude area dealing with war and punishment. However, the main reason why these actual attitudes were not "measured" in the present study was that it would have made it difficult, if not impossible, to obtain the cooperation of suitable populations of *Ss*, but other reasons did exist.

An investigation of the actual attitudes of these groups would have to be set up in multidimensional terms, not only to be comparable to the analysis of perceived attitudes, but also because some important differences between groups are probably reflected in relationships among dimensions rather than in responses to single scales. On the surface, then, the proper approach would appear to be to obtain responses to the three Thurstone scales, intercorrelate items, and factor-analyze. However, the vector model of factor analysis and the interpoint distance model of multidimensional scaling are not directly comparable (6), so the actual and perceived attitude structures could not be compared anyway.

RESULTS[8]

Each attitude booklet contained responses to 210 AB statement combinations and to 210 BA combinations. Each response, which was signified by a cross in one of a series of 11 boxes, indicated the judged extent of the imaginary other person's agreement with Statement B. This set of responses, 420 per booklet, constituted the raw data, which were analyzed separately for the seminary and Air Force groups.

The Seminary Group Data

The 420 combinations of statements were scaled by a graphical successive-intervals technique developed by Gertrude W. Diederich and described in (6). The 210 scale values for the AB statement-pairs represented the 210 possible distances among the 21 statements, all distances being estimated in one specified direction. The scale values for the 210 BA statement pairs represented the same distances, estimated in the opposite direction. There was adequate agreement between corresponding AB and BA values, so the two estimates were averaged for each distance. The general multidimensional scaling procedure described by Messick and Abelson (7) was then used to obtain a multidimensional configuration for these interstatement distances.

The end product of the multidimensional scaling procedure is a matrix of projections, which is analogous to the factor matrix of

[8] The data, along with a complete description of analytical procedures, appear in the writer's dissertation (6).

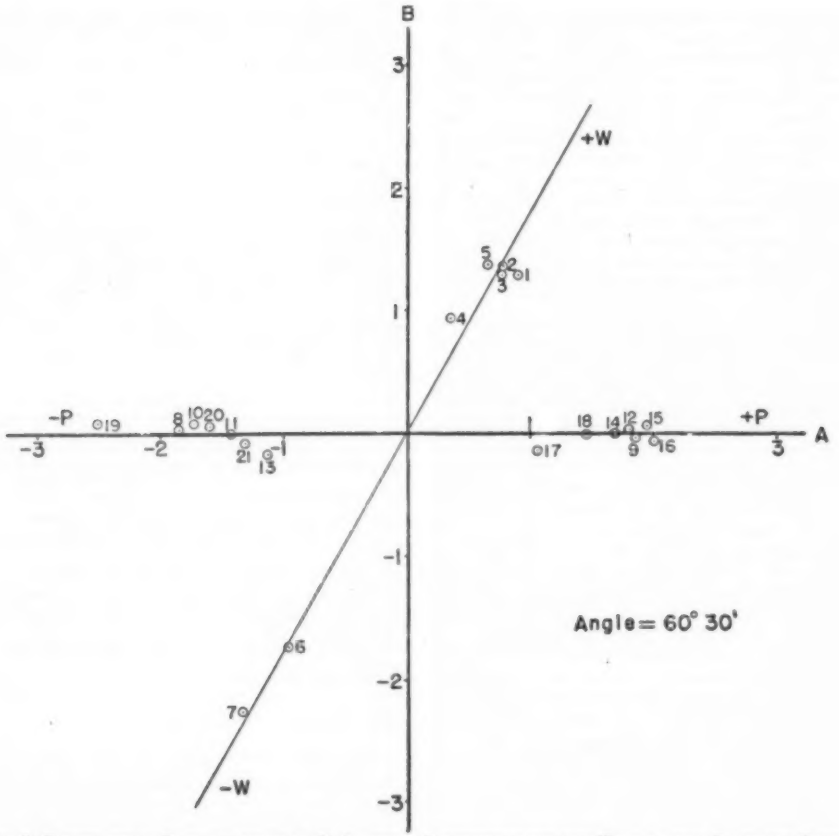


FIG. 5. DIMENSIONAL CONFIGURATION OF 21 ATTITUDE STATEMENTS FOR THE THEOLOGICAL SEMINARY GROUP

multiple-factor analysis. The rank of this matrix is equal to the dimensionality, r, of the Euclidean space defined by the experimentally obtained distances, and the elements of the matrix represent the projections of the statements on a set of r orthogonal axes placed at the centroid of the statements (13, 18). In the present experiment, this matrix of projections was of Rank 2, and when the two-dimensional configuration was plotted, two distinct streaks of points were evident. The structure was translated slightly in order to place the origin at the intersection of the streaks and then was rotated orthogonally until the largest streak of points was in a horizontal position. This particular orientation has the advantage of producing the maximum number of zero projections for this two-dimensional space, thus permitting a simple psychological interpretation. The rotated, translated configuration obtained from the seminary group appears in Fig. 5.

The horizontal dimension in this figure contains all 14 statements from the scales of attitude toward the treatment of criminals and attitude toward capital punishment, and thus it represents a "propunishment vs. antipunishment" dimension. Statements 9, 12, 14, 15, 16, and to a slightly lesser extent 17 and 18 define the propunishment end of the continuum, and Statements 8, 10, 11, 13, 19, 20, and 21 define the antipunishment end. A "war" dimension is located oblique to the "punishment" dimension at an angle of approximately 60°, with Statements 1, 2, 3, 4, and 5 lying in the prowar direction and Statements 6 and 7 in the antiwar direction. The 60° angle between the dimensions indicates a positive correlation between

the prowar and propunishment attitudes perceived by the seminary group.

The Air Force Group Data

The graphical scaling technique used to analyze the seminary data is a simple, quick iterative routine which yields a very good approximation to a computationally laborious least-squares solution for successive intervals, which was also developed by Diederich (see 6). After the seminary data had already been analyzed graphically, an IBM routine for the computational procedure was developed, so the Air Force data were treated according to the punched-card technique (15). When the 420 scale values obtained in this way were examined, it was again found that there was adequate agreement between corresponding AB and BA values, so the two estimates were averaged for each distance. The resulting 210 averages represented all possible distances among the 21 statements. The general multidimensional scaling technique (7) was used to obtain the configuration of the statements, and the matrix of projections was again found to be of Rank 2.

Two distinct streaks of points were again evident in the two-dimensional configuration, and the structure was translated slightly and rotated orthogonally to the same orientation as for the seminary data. This rotated, translated configuration for the Air Force group

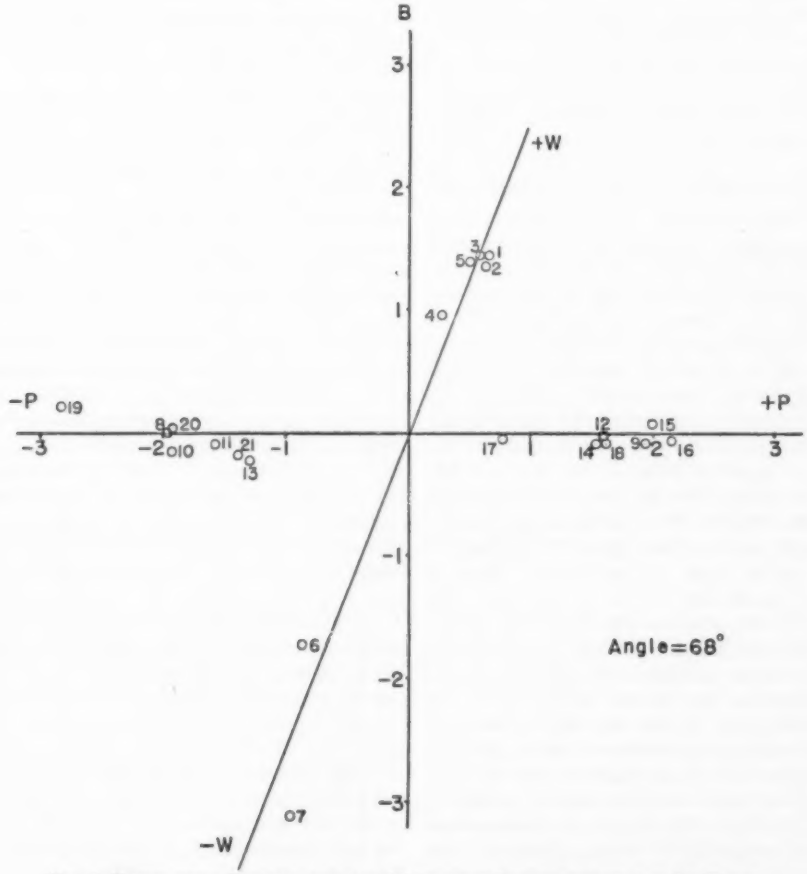


FIG. 6. DIMENSIONAL CONFIGURATION OF 21 ATTITUDE STATEMENTS FOR THE AIR FORCE OFFICER CANDIDATE GROUP

appears in Fig. 6. The horizontal factor in this figure represents a punishment dimension, having Statements 9, 12, 14, 15, 16, 17, and 18 in the propunishment direction and Statements 8, 10, 11, 13, 19, 20, and 21 in the antipunishment direction. A war dimension is located oblique to the punishment dimension at an angle of approximately 68°, with Statements 1, 2, 3, 4, and 5 lying in the prowar direction and Statements 6 and 7 in the antiwar direction. The 68° angle between these dimensions indicates a positive correlation between perceived prowar and propunishment attitudes for the Air Force group.

The two configurations described above can be considered to be adequate dimensional representations of the experimentally obtained distances, since, apart from small residuals, only two dimensions were necessary to account for relationships among 21 stimuli, and no indications of non-Euclidean properties were found, i.e., there were no markedly negative latent roots.

A Comparison of the Structures Obtained from the Seminary and Air Force Groups

A comparison between the final rotated structures obtained from the seminary and Air Force groups (Fig. 5 vs. Fig. 6) indicates an extremely high amount of agreement. It is quite evident from these figures that the two structures are essentially identical. The angle between the two dimensions in the seminary structure is 60°, while in the Air Force structure the angle is 68°. This shift in axis position is in an expected direction in terms of the probable differences between the "actual" attitudes of the Air Force and seminary groups. One might expect a theological seminary population to be more "humanitarian" than a military population. If this is the case, the "actual" attitudes of the seminary group would be expected to exhibit a higher correlation between antiwar and antipunishment attitudes than the "actual" attitudes of the Air Force group. If the "actual" attitudes of the *S*s in this particular experimental situation affect attitudes as perceived in others, then the shift in axes toward a higher correlation for the seminary than for the Air Force group is in an expected direction. However, the angular difference of 8° is probably not significant.

DISCUSSION

The dimensional configurations obtained from both the seminary and Air Force groups certainly demonstrate that perceived relationships among attitudes toward war, capital punishment, and the treatment of criminals can be adequately represented in dimensional terms. Apparently, then, individuals do think of these attitudes as having a definite structure, and when asked to make judgments concerning attitude statements, they respond in terms of such dimensional frames of reference. The configuration obtained from both the Air Force and seminary groups is essentially the same as Type B given in Fig. 2, with the war dimension drawn obliquely to the punishment dimension. Within this frame of reference, the two groups find it possible to conceive of people in each of the four quadrants, but it should be noted that certain kinds of people would be almost inconceivable to them. For instance, both groups would have difficulty in conceiving of a person who was in favor of capital punishment of extreme criminals and at the same time against punishing minor criminals.

The finding of an underlying structure in an attitude domain, when considered along with some of the specific properties of the structure, has important implications for attitude measurement and theory. Since one of the more reasonable ways to measure multidimensional attributes is to apply appropriate unidimensional scales along each of the dimensions involved, dimensional configurations such as those found in the present experiment would aid considerably, first of all, in determining the number and the nature of the relevant dimensions and, secondly, in selecting and evaluating the appropriate unidimensional scales. It is striking that the attitude statements in both the Air Force and seminary structures form straight lines, indicating a unidimensionality in the original Thurstone scales, at least for the items considered in the present experiment. Similar studies investigating the linearity of other scales would be interesting and very important in justifying the use of these scales in attitude measurement. Another important aspect of these configurations is the fact that the scales on attitude toward capital punishment and attitude toward the treatment of criminals, which are admittedly similar in content but nevertheless had been

independently established by equal-appearing interval techniques, should be perceived as precisely the same unidimensional continuum. This finding has important theoretical implications, since it suggests that different attitude variables, which had been separately abstracted on a priori grounds, can be reasonably thought of in terms of a single dimension. These results, however, pertain only to "perceived" attitudes and, no matter how suggestive they are, the structure of "actual" attitudes should be investigated before such a finding is incorporated into attitude theory. It well may be, for instance, that the "actual" attitudes of the *S*s in the present experiment were three-dimensional, even though the groups *perceived* the attitudes in others in terms of a two-dimensional configuration. In other words, the results of the present study are immediately applicable to considerations of the ways in which individuals conceive of attitudes as being structured "out there," but further research is necessary before they may be applied to considerations of individuals' own attitudes. In any event, multidimensional analysis applied in attempts at a systematic understanding of the psychology of attitudes offers the hope that the host of trait names and labels in the attitude area can be represented in terms of a limited number of variables.

The high correlation between the configurations obtained from the Air Force and seminary groups indicates a definite similarity in the ways in which these groups perceive attitudes in this particular domain as being structured. It is not unreasonable that perceived attitude structures should be similar for two diverse groups, since similar frames of reference are important for efficient communication, but the essential identity of the two structures requires comment.

Previous studies (5, 14, 16) have shown a persistent correlation between a person's own attitudes and his estimates of the attitudes of others. If the seminary and Air Force groups differ, as they probably do, with respect to attitudes toward war and punishment, the seemingly projective nature of the task of estimating the opinions of others would lead one to expect a difference between the structures obtained in the multidimensional analysis. However, the above studies are not really comparable to the present experiment, since they involved estimates of group opinion obtained from subjects who were themselves members of the group. Also, as Campbell (2, p. 23) has pointed out, "... in a group of any size, the respondent has an uneven acquaintanceship, and probably associates more with those of tastes like his own. Basing his estimates of group opinion upon his own experience in the group, his error may be in part a 'sampling bias' as well as biased perception."

The information supplied the subject in the present experiment was such as to be favorable to objective perception and unfavorable to projection. In one sense the information supplied was minimal; the other person, whose attitude relationships were to be estimated, was characterized only by complete agreement with a single statement. In the absence of a detailed characterization of the other person, the *S* evidently ascribed to his attitudes some kind of modal or average structure. It might be argued that minimal information is a favorable condition for eliciting projective responses, but the information supplied in the present case, although minimal in characterizing the other person, was quite specific concerning isolated aspects of his attitudes. A great deal of specific information is given to the subject when he is told that another person strongly agrees with the statement, "There is no conceivable justification for war," and it is not difficult to evaluate objectively the extent of this person's agreement with the statement, "On the whole, wars do the world some good." Projection is favored by conditions of insufficient or conflicting information, but when specific and relevant information is available concerning a particular judgment, an objective response is probably favored.

SUMMARY

The primary purpose of this study was to see whether a set of perceived attitude relationships could be adequately represented in dimensional terms and, if so, to discover the number and the nature of the relevant dimensions involved. A secondary purpose was to investigate judgments made by two diverse groups in order to discover whether they perceive attitudes as being structured in different ways. Accordingly, a multidimensional method of successive intervals, which is a Euclidean distance model, was applied to judgments of

attitude relationships. The procedure required *S* to arrange all possible pairs of a set of attitude statements on a distance continuum according to the degree of similarity of the members of each pair.

Seven statements were selected from each of the three Thurstone scales on attitude toward war, capital punishment, and the treatment of criminals, and these 21 statements were set up in booklet form for group presentation. The attitude booklets were administered to 40 third-year, male students at a theological seminary, and to 82 male Air Force officer candidates.

The multidimensional attitude structures perceived by the two diverse groups were essentially identical, the relationships among statements of the three attitudes being adequately represented in terms of two oblique dimensions, a war dimension and a punishment dimension. Apparently individuals do perceive attitudes, at least in this particular attitude area, in terms of a definite structure, and when called upon to make judgments concerning attitude relationships, they respond in terms of this dimensional frame of reference.

REFERENCES

1. ABELSON, R. P. A technique and a model for multidimensional attitude scaling. *Publ. Opin. Quart.*, in press.
2. CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15–38.
3. GULLIKSEN, H. A least squares solution for successive intervals assuming unequal standard deviations. *Psychometrika*, 1954, **19**, 117–139.
4. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, **9**, 139–150.
5. KATZ, M. R. A hypothesis on anti-Negro prejudice. *Amer. J. Sociol.*, 1947, **53**, 100–104.
6. MESSICK, S. J. The perception of attitude relationships: a multidimensional scaling approach to the structuring of social attitudes. Unpublished doctoral dissertation, Princeton Univer., 1954. Also, Princeton: Educational Testing Service, 1954.
7. MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, in press.
8. PETERSON, R. C. *Scale for the measurement of attitude toward capital punishment.* Chicago: Univer. of Chicago Press, 1931.
9. PETERSON, R. C. *Scale for the measurement of attitude toward war.* Chicago: Univer. of Chicago Press, 1931.
10. RICHARDSON, M. W. Multidimensional psychophysics. *Psychol. Bull.*, 1938, **35**, 659–660.
11. SAFFIR, M. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 179–198.
12. TORGERSON, W. S. A theoretical and empirical investigation of multidimensional scaling. Unpublished doctoral dissertation, Princeton Univer., 1951.
13. TORGERSON, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17**, 401–419.
14. TRAVERS, R. M. W. A study in judging the opinions of groups. *Arch. Psychol.*, N. Y. No. 266, 1941.
15. TUCKER, L. R., MESSICK, S., & GARRISON, H. A punched card procedure for the method of successive intervals. Princeton: Educational Testing Service, 1954.
16. WALLEN, R. Individual's estimates of group opinion. *J. soc. Psychol.*, 1943, **17**, 269–274.
17. WANG, C. K. A., & THURSTONE, L. L. *Scale for the measurement of attitude toward the treatment of criminals.* Chicago: Univer. of Chicago Press, 1931.
18. YOUNG, G., & HOUSEHOLDER, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 1938, **3**, 19–22.

Received January 23, 1955.

MEMBERS' REACTIONS TO APPARENT GROUP APPROVAL OF A COUNTERNORM COMMUNICATION

HAROLD H. KELLEY AND CHRISTINE LIPPS WOODRUFF[1]

Yale University

WHAT happens if, while members of a group are listening to a persuasive communication contrary to their group norms, they learn that other members have expressed approval of the communication? This question is of theoretical interest for two somewhat different reasons.

First, various theories suggest that one way to change norm-related behavior or opinions is to communicate information indicating that the norms are different from what the audience had previously believed them to be. Lewin's formulation of this idea, which was at the basis of the research on the group decision method of change, is as follows: "If the resistance to change depends partly on the value which the group standard has for the individual, the resistance to change should diminish if one . . . changes the level perceived by the individual as having social value" (5, pp. 471–472). Whereas the group decision method attempts to bring about an actual change in group consensus by means of a face-to-face discussion, a different approach to the same general problem has been provided by investigations of the effects of "majority opinion" (e.g., 1, 2, 3). In essence, these studies have attempted to change perceptions of norms by means of simple one-way communications that convey alleged or real information about the consensus of opinion among various classes of people. A variety of different kinds of information about majority opinion has been investigated and in most cases the information has been found to be effective in influencing the recipients' opinions.

The question raised in the present study lies at the intersection of these different lines of research. Can the norm-related opinions of the members of a relatively small face-to-face group be modified without discussion-mediated changes in the consensus, but merely by the transmission of new information about it? The present interest is in a procedure for transmitting information suggesting that other members of the group have approved a communication contrary to its norms. Evidence from the above mentioned research suggests the following hypothesis about this procedure: If the information can be conveyed in a credible manner, members who receive it during exposure to the counternorm communication will be more influenced by the communication than members receiving no such information during exposure.

The second reason for interest in the described procedure revolves around the problem of the credibility of the new information. This is obviously important when dealing with opinions within small, fairly intimate groups where each member is likely to have reasonably definite ideas about his colleagues' opinions on various issues. If the new information is to facilitate opinion change by modifying existing beliefs about group opinions or norms, it must be incompatible with these beliefs. Can we, then, expect the new information to be accepted? Will members of our audience believe that the consensus is changing and change their own opinions accordingly? Or will their prior belief about the norms tend to resist change, and if so, what form will this resistance take?

A number of studies have been made of similar situations in which the recipients of a persuasive communication find its contents to be incompatible with what they believe to be the attitudes of its source (the person or persons endorsing or presenting it). As summarized elsewhere (4, pp. 40–48), this kind of situation has a number of different effects of

[1] Mrs. Woodruff is now at the University of Pittsburgh and Kelley is presently at the University of Minnesota. The authors are greatly indebted to the students and faculty at Willimantic State Teachers College, Willimantic, Connecticut. Without their generous and intelligent cooperation, this investigation would not have been possible. We are also grateful for Carl I. Hovland's suggestions on the manuscript and for the able research assistance provided at different stages by Dorothy D. Brown, Thomas R. Kemm, Robert H. Peters, and Neal Woodruff, Jr. This study was conducted as part of the Yale Communication Research Program which is supported by a grant from the Rockefeller Foundation and is under the general direction of Professor Hovland.

which acceptance of the communicated viewpoint is only one. Other effects include (*a*) denial of the source's responsibility for the communication or attribution of special motives to him, (*b*) reinterpretation or distortion of the content of the message so as to make it less incongruous with beliefs about the source, and (*c*) changes in attitudes toward the source. To date, communication researchers have been interested mainly in the degree to which the communication content is accepted, and hence have planned their procedures so as to minimize these other effects. In contrast, one of the major purposes of the present study was to create a situation where some of these effects would occur frequently and where their occurrence in relation to acceptance of the communication could be examined systematically.

METHOD

The research was conducted at a small teachers' college believed to represent a rather "modern" point of view with respect to such teaching problems as discipline, evaluating pupil progress, teacher-pupil relations, learning through activities and projects, and integration of the various subjects. To measure the attitudes related to this supposed norm, a preliminary questionnaire was constructed on the basis of interviews with students and faculty. It consisted of 31 opinion items of the Likert type, including such statements as: "It is more important that children first think and attack problems than that they first learn required content" and "A standardized curriculum adequately meets the needs of the greatest number of children."

To assure ourselves that a norm related to these questions existed at this college, preliminary surveys were conducted. The questionnaire was given in 1952 to 52 freshmen, 46 juniors, and 29 seniors; in 1954, to 55 freshmen and 44 juniors. These samples included 80 to 90 per cent of the total membership of each class. On both occasions the juniors and seniors were found to exhibit greater favorability toward "modern" teaching practices than the freshmen. The p values for both the freshman-junior and freshman-senior comparisons in 1952 were less than .001. For the freshman-junior comparison in 1954, the p value was approximately .20. The relative smallness of the last difference was in part attributable to the fact that the opinions were measured late in the school year by which time the freshmen had apparently absorbed some of the prevailing ideas of the college group. The gain in favorability toward "modern" teaching practices made by the junior class of 1954 was clearly revealed in a comparison of their 1954 scores with scores they made as freshmen in 1952: for 34 students who were questioned on both occasions, $p < .001$. In short, both cross-sectional and longitudinal analyses of students at this college indicated that they acquire more positive attitudes toward modern methods of teaching as they progress through the training program. This provided some justification for the assumption that a group norm favoring modern practices exists in this college community.

The experimental procedure was carried out with volunteer members of the 1952 and 1954 junior classes who were asked to act as judges in a study of speaking effectiveness. Two to four *S*s were present at each session of the experiment, but they were seated so they could not observe one another and could react to the communication independently. They were asked to listen to a recorded speech, supposedly for the purpose of making judgments of the speaker, his delivery, and the clarity of his arguments. The speech, attributed to a professor of education from a midwestern university, called for a return to more traditional classroom methods than those advocated by so-called "progressive" educators. The major argument was that recent research on learning and motivation had cast serious doubt upon the validity of certain assumptions underlying modern teaching procedures. Particularly emphasized was the view that in order to prepare children for the demands of adult life, the schools should insist on greater respect for the authority figures in society, the fulfillment of standard requirements, and the acquisition of basic knowledge and skills.

Seven times during the ten-minute recording of the speech, one of the speaker's major points was followed by applause (clapping of hands) apparently made by the audience present at the time the speech was recorded. The experimental manipulation consisted of identifying the applauding audience differently for different *S*s. Those at certain experimental sessions were told that the speech had been recorded at a meeting attended solely by faculty members and recent graduates of their college. The *S*s at other sessions were told that the audience present at the recording of the speech was a group in a neighboring city, composed of college-trained people interested in community problems related to education. In view of these preinstructions, *S*s in the former variation (to be referred to as the "*members' applause*" variation) were confronted with a situation in which prestigeful *members* of their college group were heard applauding a speech directly opposed to the norms of the college. On the other hand, *S*s in the latter variation (to be referred to as the "*outsiders' applause*" variation) heard the identical speech being applauded by a group of other people, comparable in education and interest to their own group but *outsiders* with respect to that group. The two kinds of instructions were given to the different experimental sessions in a prearranged sequence, the basic order being an alternating one with several extra sessions being selected at random to receive the *members' applause* instructions, in order to provide a larger N for this condition. Thirty-eight *S*s (23 in 1952 and 15 in 1954) were given the *members' applause* version of the preinstructions, and 23 *S*s (14 in 1952 and 9 in 1954) were given the *outsiders' applause* version. (The 1952 and 1954 data were very similar and, consequently, were combined.)

Measures of Possible Effects

To assess systematically a number of the possible outcomes of this situation, measurements of the following variables were undertaken:

1. Changes in opinions about educational practices. Immediately before and after exposure to the communi-

cation, *S*s were given a number of items (15 before and 14 different ones after) from the preliminary questionnaire. These were presented as having no relation to the present study of speaking effectiveness but as part of an attempt to standardize a questionnaire on teaching methods for use in various teachers' colleges. At the time the second half was given, the double administration was explained as necessary to determine split-half reliability. The responses to these items, scored by the Likert method, were combined for each *S* to yield his "before" and "after" scores on opinions about teaching methods. The difference between "before" and "after" scores was taken as a measure of change in these opinions.

2. Changes in perception of the consensus of opinion within the college. After the experiment the *S*s were asked to estimate how the other people in the professional training program at the college would answer 13 of the opinion questions used above. These estimates were combined for each *S* to yield an indication of how modern a viewpoint he believed the college group generally to hold. For most of the *S*s, similar data were available from the preliminary questionnaire which had been administered several weeks before the experiment.

3. Distortions in judgments of the communicator. After the experiment the *S*s also estimated how the speaker would answer the 13 opinion questions. The purpose of this was to obtain a simple, objective measure of the tendency to distort or to misinterpret the speech, the assumption being that the opinions *S*s attributed to the speaker would reflect their understanding of what he had said in the speech. This assumption was borne out by the relation between the opinions *S*s attributed to the speaker and the accuracy with which they were able, at the end of the experiment, to summarize four of the major points of the speech. In virtually all cases in which these summaries were inaccurate, they depicted the speech as being more favorable toward modern educational practices than it actually was. The *S*s who gave distorted summaries of the speech tended consistently to attribute more "modern" attitudes to the speaker than did *S*s who gave accurate summaries (difference significant at well beyond the .001 level). Because of the objectivity with which the estimates of the speaker's opinions could be scored, they were used as the sole measure of the tendency to distort the communication content.

4. Interpretations of audience applause for the communication. At the end of the experiment the *S*s were asked open-ended questions about (*a*) whether the applause interfered with their paying attention to the speaker, (*b*) what they thought of the applause, and (*c*) why they thought the audience applauded the speaker. These questions were included to obtain an indication of the extent to which *S*s tended to deny the reality of the applause by the alleged members of the audience or to doubt their sincerity in applauding.

The reader will note that all of the above measures of effects are based on data obtained immediately after the communication situation, so the results will indicate only the *immediate* effects of this procedure. Shortly after the experiment *S*s were informed of its purpose and the details of the procedure. No subsequent measures of effects were obtained.

RESULTS

Did the *members' applause S*s, as a result of hearing the college audience applaud the counternorm speech, change their perceptions of the college norms about teaching practices? Apparently not. In the postexperiment estimates of the college opinions on this issue, there is virtually no difference between the *members' applause* and *outsiders' applause S*s. Neither is there a significant difference between the two variations when account is taken of changes in these estimates from the early preliminary questionnaire.

There is, however, support for the hypothesis that the *members' applause S*s would show greater change in their own opinions, in the direction of the speech, than the *outsiders' applause S*s. (The groups were comparable in their "before" scores, so we present only the results of the analysis of their change scores.) A t test of the significance of the mean difference in amount of change yields a p value of approximately .09. This test may not be entirely justified because the variability in amount of change is greater (although not significantly so) for the *members' applause S*s than for the other variation. Examining the two distributions of amount of change, it is apparent that they overlap at the lower values; the difference is that the *members' applause* distribution includes a higher proportion of large change scores. This difference is most marked when the median amount of change for all *S*s is taken as a cut-off point: 63 per cent of the *members' applause S*s exhibit more than this amount of change whereas only 30 per cent of the *outsiders' applause S*s do so ($p < .02$). It appears, then, that some of the *members' applause S*s were considerably affected in their opinions by the communication while others were no more affected than the *outsiders' applause S*s.

On the basis of the earlier discussion, it would seem reasonable to suppose that the occurrence of opinion change for the *members' applause S*s would depend upon their changing their perceptions of the college norms. This would not be expected for the *outsiders' applause S*s for whom the applause contains no information relevant to the college norms. Accordingly we would expect the correlation between amount of opinion change and the extent to which the college group is viewed as holding less "modern" views than judged

earlier (as measured by the preliminary questionnaire given some weeks before the experiment) to have a higher positive value for the *members' applause* *S*s than for the other *S*s. The obtained correlations tend to conform with this supposition but are not different enough in magnitude to give us much confidence in the result. The product-moment correlation coefficient is +.31 for the 32 *members' applause* *S*s from whom both sets of data were obtained as compared with the correlation of −.06 obtained from 21 *outsiders' applause* *S*s (p value for difference = .21).

Is there any difference between the *S*s in the two experimental conditions in their tendency to distort the communication? As reflected in their judgments of the speaker's answers to a series of opinion questions on educational practices, the *members' applause* *S*s clearly exhibit more distortion than do the other *S*s. All *S*s judge the college group to have quite modern views on these questions and they judge the speaker to hold a much less modern view. However, the *members' applause* *S*s judge him to be more modern than do the *outsiders' applause* *S*s, a t test for the significance of the mean difference yielding a p value of about .015. As in the case of the distributions of opinion change, the *members' applause* *S*s show greater variability in the opinions they attribute to the speaker (p = approx. .01), which makes the interpretation of the t test somewhat ambiguous. Once again, the difference between the two variations is shown most strikingly by a simple comparison of the proportion of *S*s in each who are above the median of their combined distribution. Sixty-five per cent of the *members' applause* *S*s as compared with only 26 per cent of the *outsiders' applause* *S*s attribute more modern attitudes to the speaker than does the median subject ($p <$ .004). It appears, then, that when the applause for the speaker is attributed to members of the college group, *S*s tend somewhat to perceive the speaker as similar to the college group in his opinions.

Finally, an analysis was made of the *S*s' postexperimental statements about their reactions to the applause. Working without knowledge of the experimental variation to which each *S* belonged, a judge evaluated these statements in terms of the degree to which each *S* seemed to believe that the applause did not indicate approval of the speaker's points but had some other significance. Fifteen *S*s suggested that the applause was "dubbed" into the recording and 25 *S*s suggested that the audience applauded only to be polite or because it was expected of them. Of the *S*s entertaining one of these hypotheses (who comprised about two-thirds of all *S*s) about half appeared to be fairly certain that their alternative explanation was correct.

Of the *S*s in the two variations, we might expect the *members' applause* *S*s more often to doubt the authenticity or sincerity of the applause, because they should find it more difficult to imagine members of their college group approving the counternorm communication. However, a comparison of the two experimental variations reveals no difference between them.

The foregoing results suggest that the *members' applause* condition produced rather varied effects—opinion change, distortion of the speaker's point of view, and (although to no greater degree than the *outsiders' applause* condition) special interpretations of the audience's applause for the speaker. These are the major kinds of responses we might expect of group members when they are confronted with the incongruous spectacle of other members of their group apparently applauding a speech that is quite contrary to their shared norms. The discrepancy between prior conceptions of group norms and this new "evidence" can be dealt with in a variety of ways: abandoning the old views of the norms (which does not occur to any great degree in the present study) and *changing* one's opinions to conform with the group-sanctioned communication, *distorting* the speech so that it conforms to one's understanding of the norms, and *dissociating* the audience from the applause by describing it as "fake" or due to special circumstances.

What is the relation among these three responses of changing, distorting, and dissociating? If they represent alternative ways in which the *members' applause* *S*s can adapt to or make sense of the incongruous situation, we might expect them to be mutually exclusive: to the extent that a person in this variation handles the situation one way, it would be less necessary for him to resort to the others. For the *outsiders' applause* *S*s on the other hand, these responses do not constitute ways

of dealing with a problem situation of the sort described. Insofar as they do occur for these *S*s, the three reactions might be determined by more or less haphazard and unrelated cues and hence might be expected to appear independently of one another.

This possibility was investigated by analyzing the interrelations among the three responses. In the *members' applause* sample, the correlations between the pairs of variables are all negative but only the coefficient obtained from the distortion-dissociation comparison differs significantly from zero ($r = -.35$, $p < .05$). Examining the two scatterplots involving opinion change, in each instance it appears that the correlation is low primarily because of a number of cases that are low on both variables. Furthermore, these cases tend predominantly to be high on the third variable, as might be expected if the three tend to be mutually exclusive in their occurrence. For example, when we consider opinion change plotted against distortion, there are a number of cases low on both variables and these tend to be cases high on dissociation. This suggests correlating opinion change with a score reflecting how much a person exhibits *both* of the other reactions. This was done by simply summing the distortion and dissociation scores for each person (after equalizing their variances). The resulting score would be significantly correlated (in the negative direction) with opinion change were it not for the presence of two cases who are low on all three scores. A chi-square test of the relationship, which uses the two medians as cutting points to produce a 2 x 2 table and is relatively insensitive to the two deviant cases, yields a value significant at the 5 per cent level. Within the *members' applause* sample, then, the results tend to be consistent with the hypothesis of mutual exclusiveness: *S*s showing one response to a marked degree tend to exhibit the others to a low degree. The data are fairly convincing with respect to the negative relationship between distortion and dissociation but are merely suggestive of a negative relation between these responses and opinion change.

Within the *outsiders' applause* sample, a different pattern of relationships prevails. None of the intercorrelations are significant nor is opinion change related to the sum of the other effects. In brief, the three responses seem to occur independently of one another.

The contrast between the sets of data from the two samples was examined quantitatively in the following manner. For each person, it was determined whether or not, for each of the three variables, he ranks in the upper third of the distribution (to be referred to as "high"). The pattern of "highs" and "lows" for each *S* was then tabulated. If the reactions occur in a completely mutually exclusive manner, we would expect one third of the *S*s to be "high" only on change, another third to be "high" only on distortion, and the remaining third to be "high" only on dissociation. On the other hand, if the three reactions occur independently of one another, we would expect $\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3}$ of our *S*s to be "high" on all three variables, $3 \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3}$ to be "high" on two of the three, etc. For the *outsiders' applause* sample, the obtained frequencies for the various patterns are fairly close to those predicted by the hypothesis of independence. (A chi-square test of goodness of fit yields a p value of approximately .85.) For the *members' applause* sample, however, the obtained frequencies deviate significantly from the expected ones ($p < .05$) and they do so in the direction of the frequencies predicted by the hypothesis of mutual exclusiveness. Although only 16.5 of the latter sample would be expected, according to the hypothesis of independence, to rank "high" with respect to one and only one response, 24 are actually found to do so ($p < .02$). Comparing the two samples, a larger percentage of the *members' applause* *S*s rank "high" with respect to one and only one of the response variables (65 per cent compared with 39 per cent, which yields a p value of .052). In short, the patterns obtained in the *outsiders' applause* sample tend to conform to the hypothesis of independent occurrence of the three responses whereas those from the *members' applause* sample deviate from this hypothesis in the direction of conformity with the hypothesis of mutual exclusiveness. Our earlier results show that the latter effect is attributable largely to the negative correlation between distortion and dissociation within the *members' applause* sample.

DISCUSSION

This study shows the effects of attempting to use a one-way communication to change perceptions of norms when the audience is composed of members of a relatively small and

tightly knit group. It should be emphasized that the results bear only upon the immediate effects of this procedure since all measures of effects were taken immediately after the communication was presented. It is very unlikely that effects based on measures taken after any considerable interval of time would be identical with those observed, especially in view of the possibility that the experimental procedure would have stimulated some of the *S*s to obtain additional direct information about the group norms.

Within the above limitations, some support is provided for the hypothesis that change of group-anchored attitudes by a counternorm communication can be facilitated by conveying information that members of the group approve the opinions advocated in the communication. Specifically, the data indicate that listeners to a recorded speech incompatible with their norms tended to accept the content of the speech to a greater extent when given evidence that other members of their group have approved the speech than when given evidence that an irrelevant group of "outsiders" have approved it.

Whether or not this greater acceptance of a counternorm point of view is mediated by changes in perceptions of the norms is not clearly indicated by the data. The *S*s who had been given the new evidence about the group's opinions showed no greater change in perceptions of the group norm than did the *S*s who received no such information. This may be indicative of the great difficulty involved in changing perceptions of norms in a group such as the one under investigation. Considering the great amount of evidence about the norms available to members before the communication, we would hardly expect them to be much affected by this new piece of information which is contradictory to what they have previously learned. Even so, the lack of difference between *members' applause* and *outsiders' applause* *S*s in their perceptions of the norms is a bit difficult to reconcile with the greater tendency for the former to change their opinions (which, by other evidence, appear to be related to group norms) and with the suggestive evidence that the amount of opinion change is correlated (within the former variation) with the amount of shift in perception of the norms. In general, the evidence on this point is quite inconclusive. The number of cases involved is probably far too small to permit a precise test of the hypothesis that opinion change is dependent upon change in perceived norms, so while the present data indicate how the hypothesis can be tested, they neither confirm nor disprove it.

The findings from this experiment also suggest some answers to the broad problem of how people react to a persuasive communication, supported by prestigeful authority figures, that conflicts with and contradicts much of their pre-existing knowledge and opinions. Three of the possible reactions to this discordant situation—opinion change, distortion, and dissociation—were identified and measured. The evidence on their interrelationships, although far from conclusive, indicates that they tend to occur in a mutually exclusive manner or, as we might say, they tend to be negatively interdependent. This suggests that if a person handles this incongruous situation by one of these "adaptations," the necessity of resorting to the others is thereby obviated.

We may speculate that the following processes occur in the case of most *S*s in the *members' applause* variation: Under the impact of the experimental procedure, they experience a certain amount of pressure to change their conceptions of the group norms and their own opinions about teaching practices. This pressure to change meets with considerable resistance. The existing conceptions of the norms, being based on a good deal of direct contact with other members of the group, are not readily overturned; the *S*s' own opinions, being anchored in the norms, and perhaps having other kinds of supports, also tend to resist change. For those *S*s whose resistance is relatively low (or whose resistance is reduced by the new "information" about opinions in the groups), the pressure to change produces an effect upon their opinions. But for *S*s whose resistance is high, no change occurs and they are caught between the opposing forces produced on the one hand, by the pressure to change and, on the other, by the resistance. This conflict of forces produces tension or feelings of discomfort which the person is motivated to reduce in some manner. Two ways of reducing the tension are most available in the present situation: (*a*) distorting the content in such a way that the communication situation

no longer "requires" change (specifically, *S*s can interpret the speaker's viewpoint as being in agreement with their own and their group's opinions), and (*b*) dissociating the audience from the applause in such a way that it no longer means that the audience approves the speech. Both of these reactions, which consist of perceptual reorganizations of the communication situation, tend to reduce the pressure to change induced by the *members' applause* situation. Therefore, an *S* needs to resort to only one of them in order to handle the conflict between the pressure to change and the resistance.

In sum, we propose the hypothesis that the three reactions constitute intersubstitutable ways of handling the pressure to change which the *members' applause* variation induces. Changing one's opinions reduces the pressure directly through conforming to it, while distortion and dissociation can be viewed as alternative ways of transforming or redirecting the pressure. Because of their interchangeability as ways of handling the communication situation, we would expect them to occur in a negatively interdependent manner.

Extending this interpretation to the *outsiders' applause Ss*, we might assume that, as compared with the *members' applause Ss*, they feel less pressure to change their opinions in the direction advocated by the speaker. Assuming further that the two samples are equally resistant to change, we would expect the *outsiders' applause Ss* to exhibit less change, and, even though they do not change, to experience little tension over the conflict between the induced pressure and their resistance. It would then seem to follow that these *S*s would resort relatively infrequently to the two ways of reducing this tension—distortion and dissociation. Furthermore, because the various reactions would occur infrequently and weakly, determined by more or less random factors, we would not expect to observe any interdependence among them.

The only fact inconsistent with the above interpretation is the evidence that the *outsiders' applause Ss* exhibit dissociation just as frequently as do the *members' applause Ss*. The applause was actually dubbed into the tape recording of the speech and it apparently afforded enough cues to this effect that *S*s in both variations frequently entertained some doubts about its genuineness. Thus, the probability of this reaction may have been so high for all *S*s that any differential tendency was washed out. Had the artificiality of the applause been less apparent, dissociation might have been resorted to more frequently by the *members' applause S*s than by the other *S*s.

The interpretations in the preceding paragraphs assume that all of the *members' applause S*s paid equal attention to the communication and, because of exposure to it, experienced considerable pressure to change their opinions. The differential effects of the *members' applause* treatment are then traced to different ways of handling this pressure. An alternative interpretation might be made in terms of differential tendencies to pay attention to and learn the communication. We might assume that for some *S*s in the *members' applause* variation (those showing a large amount of opinion change), the applause attributed to their college group acted mainly to motivate them to pay close attention to the speech and hence to be exposed to the arguments it contained. Consequently, they would have been influenced by the communication and the extent of this influence need not have depended in any way upon their accepting the applause as indicative of a changing group norm. Other *S*s (those showing large amounts of distortion) may have been distracted by the applause, finding that it interfered with listening to the speech. We would not expect these latter *S*s to acquire a very clear impression of the speech and, when asked later about the speaker's point of view, they might have had little else on which to base their judgments than their recollection of the applause. Remembering that their group had applauded the speaker, these *S*s might reasonably have tended to attribute their group's attitudes to him.

The above interpretation in terms of differential effects of the applause upon attention is consistent with the distortion and opinion change results and with the relative lack of dependence of opinion change upon changes in perceptions of the norms. However, the authors doubt the validity of this interpretation, largely on the grounds that the experimental procedure made it unlikely that there would be marked differences in attention among the *S*s. Considering that only a small number of

*S*s were present at each experimental session and that they were instructed to listen carefully to the recording, it seems unlikely that any *S*'s attention could have been so inadequate as to account for some of the observed tendencies to misinterpret the content of the speech. We believe it more likely that this distortion was motivated, as suggested above, by a desire to avoid conflict between, on the one hand, the induced pressure to change, and on the other, the resistance to change of the relevant opinions. In short, we believe these results should be generalized not in terms of interferences with attention, but rather in terms of avoiding conflict between pressures to change and resistance against change. If our preferred interpretation is correct, the results suggest that in persuasion situations where pressure to change encounters resistance, it is likely that some *S*s will exhibit change, others will exhibit distortion of content, and others dissociation of communicator (or endorser) from the content. If, as we believe, these reactions constitute alternative ways of handling the conflict between pressure and resistance (and, of course, there are undoubtedly other ways, not revealed in this investigation), they will tend to occur in the mutually exclusive pattern suggested by the results of this experiment.

SUMMARY

In a study of the use of one-way communications to change group-anchored attitudes, members of a college group listened to a recorded speech contrary to their norms. An audience that was heard to applaud the main points in the speech was described for some subjects (the *members' applause Ss*) as composed of members of the college group; for others (the *outsiders' applause Ss*) it was described as composed of anonymous outsiders. The reactions of the *S*s were measured immediately after the speech, so the following general findings apply only to the immediate effects of this procedure.

In line with the hypothesis that the change of group-anchored opinions can be facilitated by conveying information that other group members have changed their opinions, the *members' applause Ss* exhibited more immediate opinion change in the direction advocated by the speech than did the other *S*s. The *members' applause S*s also showed a greater tendency to misinterpret the speaker's point of view, judging it to be more similar to the college norms than did the *outsiders' applause Ss*.

Some *S*s in both variations reacted to the communication by dissociating the alleged audience from the applause; they described the applause as false or attributed it to factors other than approval of the speech. For the *members' applause S*s, there is suggestive evidence that the three reactions of opinion change, distortion, and dissociation tended to be mutually exclusive—if one occurred, the others tended not to. The hypothesis is proposed that these reactions constitute alternative, intersubstitutable ways of dealing with persuasive communications that confront the individual with arguments or information, sanctioned by prestige figures, which are inconsistent with his preconceptions and prior opinions.

REFERENCES

1. ALLARD, W. A test of propaganda values in public opinion surveys. *Soc. Forces*, 1941, **20**, 206–213.
2. ASCH, S. E. Studies in the principles of judgments and attitudes: II. Determination of judgments by group and ego standards. *J. soc. Psychol.*, 1940, **12**, 433–465.
3. BURTT, H. E., & FALKENBERG, D. R., JR. The influence of majority and expert opinion on religious attitudes. *J. soc. Psychol.*, 1941, **14**, 269–278.
4. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
5. LEWIN, K. Group decision and social change. In G. E. Swanson, T. M. Newcomb, and E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952. Pp. 459–473.

Received January 18, 1955.

SOCIAL SENSITIVITY IN THE SMALL TASK-ORIENTED GROUP

J. RICHARD SUCHMAN

Department of Child Development and Family Relationships, Cornell University[1]

SOCIAL sensitivity has received increased attention from investigators in clinical, social, and educational psychology. They have recognized for some time that patterns of thought and feeling are transmitted from one person to another during social interaction, in a manner quite outside the realm of direct verbal or written communication. Therefore, it is felt that increased knowledge of the determinants of social sensitivity in the primary group will lead to a clearer understanding of group processes. The present investigation was designed to determine whether certain factors in the social situation are responsible for the level of sensitivity in any given group.

Social sensitivity is defined herein as *the perception in a social situation of those thoughts and feelings of others that relate to the situation.* In other words, the socially sensitive person perceives with accuracy the responses of others to the various stimuli in the situation, or to the entire situation as a whole.

In any given instance where the communication of thoughts and feelings is being measured, there are at least five distinct elements. Each of these is a source of variables that might determine the quality of that communication. These elements are:

1. the individual who is perceiving the feelings of another (herein designated as the estimator);

2. the individual whose feelings are being perceived by another (herein designated as the estimatee);

3. the relationship between the two individuals involved;

4. the physical, social, and psychological atmosphere in which this communication is taking place (herein designated as the group situation);

5. the nature of the estimation task (i.e., the kinds of feelings and thoughts the subjects are asked to estimate).

Since the degree of social sensitivity in any situation may be affected by variables stemming from each of these sources, the following hypotheses were set forth to be tested:

1. that individuals differ with respect to their average accuracy in estimating the feelings of others;

2. that individuals differ with respect to the average accuracy with which others can estimate their feelings;

3. that estimation accuracy is a function of the particular combination of estimator and estimatee;

4. that groups of individuals differ with respect to their total accuracy in estimating each other's feelings within the group;

5. that sensitivity varies with respect to the kinds of feelings being estimated;

6. that both groups and individuals vary with respect to which feelings they can estimate more accurately.

Six additional hypotheses were made regarding the effect of sentiment structure upon accuracy:

7. that the average accuracy of an individual is related positively to his self-favorability (self-confidence or self-approval);

8. that the average accuracy of an individual is related positively to his other-favorability (the degree to which he regards others favorably);

9. that the level of accuracy between any two members of a group is related positively to the degree to which the pair regard each other favorably;

10. that the level of accuracy between any two members of a group is related positively to the similarity between them in their feelings about themselves and toward each other;

11. that the average accuracy of a group as a whole is related positively to the average self-favorability and other-favorability of its members;

12. that the average accuracy of a group as a whole is related positively to the similarity among the members in their feelings about themselves and toward each other.

[1] The author is presently at the University of Illinois.

METHOD

Subjects

The subjects (*S*s) were obtained through an arrangement with the Octagon Club of Cornell University. This organization produces an original musical comedy each spring, and had planned to conduct a series of dramatic tryouts to appraise the available talent on campus. It was agreed that the experiment would be presented to the *S*s ostensibly as a "new type of audition."

The *S*s were assigned to three groups of five persons each. One group was entirely male, one was female, and one was mixed. Care was taken to arrange the groups so that the members did not know one another.

All of the *S*s were undergraduates at Cornell University. The age of the students ranged from 17 to 24, with an average age of 19.5 years.

Procedure

The groups met on different evenings and were subjected to identical experimental conditions insofar as this was possible. Each group was instructed to plan a short skit. Thirty minutes were allotted for this task. At the end of this planning period, social sensitivity questionnaires were distributed. After these had been filled out, the group presented their skit.

Questionnaires

The instrument used for measuring social sensitivity in the experimental groups was of the multiple-choice type. It consisted of four parts. The first two were used to establish the *criteria* (the feelings of the *S*s), and the second two were records of the *estimations* of these criteria.

Part I consisted of the following eight statements representing feelings an *S* might have had during the planning session:

1. I felt self-confident.
2. I was satisfied with what we accomplished.
3. I was fascinated by the whole thing.
4. I felt relaxed.
5. I felt left out.
6. I was frustrated.
7. I was self-conscious about being observed.
8. I felt panicky about finishing on time.

The *S*s were instructed to check one of the four alternatives after each item, indicating the degree to which the statement represented their own feelings (i.e., "very much," "a good bit," "only slightly," or "not at all").

Part II of the questionnaire consisted of eight statements designed to record feelings that an *S* might have had toward the other group members during the planning session:

1. I admired his ability.
2. I liked him as a person.
3. I felt indifferent toward him.
4. I was irritated by some of the things he said.
5. I felt pushed around by him.
6. I grew tired of him.
7. I felt close to him.
8. I was amused by what he said.

Every *S* checked a set of these items for each of the other group members. The same four-point scale was used.

Parts III and IV of the questionnaire were similar to Parts I and II, respectively. In Part III, the *S* was asked to estimate the feelings of the other four *S*s by responding to the eight statements from Part I in the way he felt each of them had responded. In Part IV, the *S* estimated how he thought each other *S* felt toward him.

For simplification, the four parts of the questionnaire may be summarized as follows:

Part I. "How did I feel?"
Part II. "How did I feel toward him?" (filled out four times, once for each other group member).
Part III. "How did he feel about himself?" (filled out four times, once for each other group member).
Part IV. "How did he feel toward me?" (filled out four times, once for each other group member).

Scoring Method

The measure of social sensitivity that was used in this investigation was an *accuracy score*. This score was a mathematical measure of an *S*s success in estimating a criterion furnished him by another *S* (i.e., how accurately A estimated the response of B on a particular item). An accuracy score was computed for each estimation made by each *S*.

As mentioned above, the *S*s were instructed to mark each criterion item "very much," "a good bit," "only slightly," or "not at all," depending upon the extent to which the *S* experienced the feeling described in the statement. In estimating how another person marked a particular item, the *S* would check one of these same four points on the scale. The following scoring system was used: A perfect estimate was given a score of 3 points; an estimate that missed the criterion by one space on the scale in either direction was given 2 points; an estimate that was off by two spaces in either direction was given 1 point; and an estimate missing the criterion by the maximum of three spaces was given a score of zero.

Total accuracy scores were then computed in the following way: It will be remembered that a given *S* estimated two sets of criterion statements for each of the other four *S*s. One set pertained to the self (*self-criterion*), and the other set pertained to the estimator (*other-criterion*). The accuracy scores for the estimations made on the self-criterion items were totaled; this was called *total self-accuracy*; it was a numerical measure of the accuracy with which A's responses to items concerning his feelings about himself were estimated by B. In the same manner, the accuracy scores on the other-criterion items were totaled and called *total other-accuracy*. This was a numerical measure of how well B could estimate the way A responded to items about A's feelings toward B.

Each of the two sets of accuracy scores in each group was arranged in the form of a 5-by-5 matrix, such as the matrix for total other-accuracy in Group A

TABLE 1

TOTAL OTHER-ACCURACY MATRIX FOR GROUP A

Estimatees	Estimators					Row total
	1	2	3	4	5	
1		20	16	20	19	75
2	18		22	21	18	79
3	19	20		18	15	72
4	21	19	19		20	79
5	20	17	21	21		79
Column total	78	76	78	80	72	384 Grand total

(Table 1). The five columns numbered across the top represent the five *estimators* (i.e., the five members of the group in their capacity as estimators). Each of these five people was at the same time being estimated by the other four. The five rows numbered on the left-hand margin of the matrix represent the same five members, but in their capacity as *estimatees* (i.e., persons whose criterion ratings were being estimated). The accuracy scores were placed in the column of the estimator and in the row of the person whose criterion ratings were being estimated. For example, in the matrix in Table 1, *S* No. 1 achieved a score of 18 when estimating the feelings of *S* No. 2. Thus, the number 18 has been entered in Column 1, Row 2. The five blank boxes that form the diagonal were not filled because the *S*s naturally could not "estimate" their own feelings.

The column totals represent the raw estimation accuracy scores for each of the five members of the group. The row totals are the total *estimatability* scores for each of the five *S*s. Thus, each *S* has one score representing the accuracy of his estimations of the others' feelings, and another score representing the accuracy with which his own responses have been estimated. The total score for the group is in the lower right corner of the matrix.

This is the basic matrix form that was used throughout the experiment. Matrices were set up in each group of *S*s for each of the 16 items, for *total self-accuracy*, *total other-accuracy*, and for the sum of the two, called *total accuracy*.

An analysis of variance design was used.[2] By combining sums of squares which were computed separately for each of the three groups, it was possible to perform a single grand analysis of variance.[3]

[2] For a discussion of the statistical problems of obtaining sums of squares from a matrix with a missing diagonal, see Henderson (5).

[3] A 19-page table containing all 57 of the matrices used in the analysis and their sums of squares has been deposited with the American Documentation Institute. Order Document No. 4652 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Project, Library of Congress.

TABLE 2

ANALYSIS OF VARIANCE OF ACCURACY SCORES FOR INDIVIDUALS

Line	Source of Variation	*df*	Mean Squares	*F* Ratio
1	Among estimators (Columns)	12	1.106	2.81*
2	Among estimatees (Rows)	12	1.195	3.04*
3	Among groups	2	2.438	6.20*
4	Among items	15	1.969	5.01*
5	Estimator-item interaction	180	.562	1.43*
6	Estimatee-item interaction	180	.559	1.42*
7	Group-item interaction	30	2.050	5.22*
8	Within-type man interaction	66	.613	1.56*
9	Average interaction over types	33	.481	1.22
10	Man-interaction discrepancy for types	33	.744	1.89*
11	Man-interaction discrepancy for items within types or error	462	.393	
12	Total	959		

* Level of significance: $p < .01$

RESULTS

Analysis of Variance

Line 1 of Table 2 represents the variance among subjects with respect to their total accuracy scores as estimators. The *F* ratio of 2.81 indicates that the differences among the *S*s in their average accuracy in estimating the responses of their fellow group members are significant ($p < .01$). This supports Hypothesis 1, "that individuals differ with respect to their average accuracy in estimating the feelings of others."

Line 2 represents the variance among the *S*s with respect to their total scores as estimatees. The *F* ratio of 3.04 indicates that the differences among the *S*s in the average accuracy with which their responses were estimated by their fellow group members are significant ($p < .01$). This supports Hypothesis 2, "that individuals differ with respect to the average accuracy with which others can estimate their feelings."

Line 3 represents the variance in total accuracy among the three groups of *S*s. The *F* ratio of 6.20 indicates that the differences among the groups are highly significant ($p < .01$). This supports Hypothesis 4, "that groups of individuals differ with respect to their total accuracy in estimating each other's feelings within the group."

Line 4 represents the variance in accuracy with respect to the 16 items of the social sensitivity instrument. The F ratio of 5.01 indicates that the differences among the items in the instrument with respect to the average accuracy achieved on each are significant ($p < .01$). This supports Hypothesis 5, "that sensitivity varies with respect to the kinds of feelings being estimated." A split-half test of reliability was made on the sixteen items of the questionnaire. The correlation between the odd- and even-numbered items was .32, corrected to .48 by the Spearman-Brown formula.

To simplify the above findings, accuracy was found to be a function of four principal factors: (*a*) who was doing the estimating, (*b*) whose feelings were being estimated, (*c*) in what group this was taking place, and (*d*) what feelings were being estimated.

Line 5 represents the variance of the interaction between estimators and items. The F ratio of 1.43 indicates that the differences among the estimators with respect to their accuracy on each of the questions is significant ($p < .01$).

Line 6 represents the variance of the interaction between estimatees and items. The F ratio of 1.42 indicates that the differences among the estimatees with respect to the accuracy with which their feelings were estimated on each of the items is significant ($p < .01$).

Line 7 represents the variance of the interaction between groups and items. The F ratio of 5.22 indicates that the differences among the three groups with respect to their accuracy in estimating responses to each of the items of the instrument were significant ($p < .01$).

The results on lines 5, 6, and 7 support Hypothesis 6, "that both groups and individuals vary with respect to which feelings they can estimate more accurately." To reduce these findings to simplest terms, it may be stated that for any given estimator, estimatee, or group as a whole, there are certain feelings that are more easily estimated than others, and that groups as well as individuals differ with respect to what these feelings are.

Line 8, the within-type man interaction, is the conventional analysis of variance interaction term. It is the interaction between the estimator and the person being estimated, with the two types of accuracy scores (i.e., total self-accuracy and total other-accuracy) treated as separate scores rather than as total accuracy. The F ratio of 1.56 indicates that the interaction between estimator and estimatee produced an effect upon accuracy that is significant ($p < .01$). In other words, over and above the average level of accuracy of a given estimator and the average level of estimatability of his estimatee, there is an additional effect upon accuracy that is a function of the unique combination of subjects.

This interaction effect has two component parts. The first of these, the average interaction over types, appears on Line 9. This represents the interaction between the estimator and the estimatee without any distinction between the two types of items. The F ratio of 1.22 indicates that this interaction is *not* significant. The second component part, the man-interaction discrepancy for types, appears on Line 10. This represents the difference in the interaction effect for any given estimator-estimatee combination that is found between accuracy scores based on self-criterion and those based upon other-criterion. The F ratio of 1.89 is significant ($p < .01$).

To restate the findings regarding the man interactions, one might say that accuracy was to some extent dependent upon whose feelings were being estimated by whom, and that this interaction effect was not the same for both types of estimation items. This supports Hypothesis 3, "that estimation accuracy is a function of the particular combination of estimator and estimatee."

In addition to the findings presented in Table 2 and discussed in the preceding paragraphs, a correlation of .652 was found between the accuracy of *S*s as estimators and their estimatability as estimatees.

Analysis of Covariance

Analysis of covariance was used to determine whether any relationship existed between accuracy and certain variables selected as measures of sentiment structure.[4] Hypotheses 7 through 12 were tested by the analysis. Hypotheses 7 and 8 pertain to the relationship between the sentiment structure of the individual and his accuracy score. Nine and 10

[4] For the purpose of covariance, total other-accuracy was used as the accuracy variable because it more nearly represented the kind of sensitivity involved in the perception of interpersonal relationships.

TABLE 3
CORRELATIONS BETWEEN ACCURACY SCORES AND SENTIMENT STRUCTURE

Measure of Sentiment Structure	Among Estimators		Within Pairs		Among Groups	
	r	p	r	p	r	p
Favorability of self-criterion ratings	.20	n.s.			.95	.20
Favorability of other-criterion ratings	.45	.10	.54	.10	.85	n.s.
Similarity of self-criterion ratings	.38	.10	.17	n.s.	.80	.05
Similarity of other-criterion ratings	.53	.01	.48	.05	1.00	.01

pertain to the mutual favorability and similarity of sentiment structure between a pair of interacting persons and the effect of these variables on accuracy. Hypotheses 11 and 12 pertain to the total group favorability and similarity of sentiment structure and their effect upon accuracy. Besides asking the question, "Does the sentiment structure of an estimator have any effect upon his accuracy?," this analysis also seeks to find out if the effect is the same for estimators in general, for pairs of *S*s estimating each other, and for groups as a whole. The correlation coefficients derived from the four analyses of covariance are presented in Table 3 along with their levels of significance.

Correlates of accuracy for the individual. Hypothesis 7 states, "that the average accuracy of an individual is related positively to his self-favorability." The estimator's self-criterion ratings were used as a measure of self-favorability. The covariance between this and accuracy yielded a correlation of .20 which is not significant and fails to support the hypothesis. Hypothesis 8 postulates "that the average accuracy of an individual is related positively to his other-favorability." The estimator's four sets of other criterion ratings were used as a measure of other favorability. These were the questionnaires on which he indicated how he felt toward each of the four other members of his group. The covariance between other-favorability and accuracy yielded a correlation of .45 ($p < .10$). This lends some support to the hypothesis that the more a person likes other people, the more accurate he will be in estimating their feelings toward him.

The covariance between accuracy and two similarity scores was analyzed. One was the similarity between the estimator's self-ratings and those of the other group members (self-criterion similarity), and the other was the similarity between the estimator's ratings of the others, and their ratings of him (other-criterion similarity). The correlations among individual estimators were .38 ($p < .10$) and .53 ($p < .01$), respectively. This means that the more similar a man's feelings are to those of the person he is estimating, the more accurate he will be. No hypotheses had been made regarding these relationships.

Correlates of accuracy within pairs. Hypothesis 9 states "that the level of accuracy between any two members of a group is related positively to the degree to which the pair regard each other favorably." In simple terms it means that the more two people like each other, the better able they are to estimate each other's feelings. The interaction effect of the covariance between other-favorability and accuracy was a measure of this relationship and yielded a correlation of .54 ($p < .10$). Here again the data appear to give some support to the hypothesis.

Hypothesis 10 states "that the level of accuracy between any two members of a group is related positively to the similarity between them in their feelings about themselves and toward each other." The interaction effect of the covariance between self-criterion similarity and accuracy had a low correlation coefficient (.17) which was nonsignificant. Where similarity of self-oriented feelings is concerned, Hypothesis 10 found no support.

The interaction correlation for the covariance between other-criterion similarity and accuracy is .48 ($p < .05$). This lends support to the part of Hypothesis 10 that postulates a positive relationship between a pair's similarity of other-criterion ratings and their combined total accuracy.

Correlates of total group accuracy. Hypothesis 11 postulates "that the average accuracy of a group as a whole is related positively to the average self-favorability and other-favorability of its members." The correlations are both high (.95 and .85), although because of the small number of groups involved, the generally accepted levels of confidence are not present. Nevertheless, since both r's are high and positive, there is reason to believe that both self-

favorability and other-favorability at the total group level do correlate positively with accuracy.

The same kinds of results were found with respect to similarity and accuracy at the group level. The correlation between similarity of self-criterion ratings and accuracy (.80, $p < .05$), and between similarity of other-criterion ratings and accuracy (1.0, $p < .01$), lend very strong support to Hypothesis 12, "that the average accuracy of a group as a whole is related positively to the similarity among the members in their feelings about themselves and toward each other."

LIMITATIONS

Brief mention should be made here concerning the principal limitations of this study before any discussion of the results is undertaken.

The sample was homogeneous with respect to age and educational level. The subjects were college undergraduates at Cornell University. In addition, they were interested in drama and in obtaining a part in a musical comedy. This sample bias must be considered as a limitation, since the extent to which these factors are related to social sensitivity is not known.

Although much care was taken to provide genuine motivation for the *S*s, the experimental situation was to some extent an artificial one. The restricting conditions under which the interaction took place were far from what one would expect to find in most real-life situations. The task of planning a ten-minute skit under observation, under the pressure of time, and with total strangers was anxiety producing and did not represent what might be considered the usual interactive situation in which social sensitivity is presumably operative.

The problem of selecting a limited number of items that would sample a relatively wide range of feelings was a difficult one. The items that were finally used were not validated as measures of genuine spontaneous feelings. It is possible that some people responded in a distorted way, feeling one way and responding in another. It must be assumed, however, that for the most part the responses of the subjects to the criterion items were reasonably honest.

Another limitation imposed by the design of the questionnaire was the fact that accuracy could be achieved in a fairly systematic fashion, with but a minimum of social sensitivity. This could have occurred when the estimator had feelings that were similar to those of the person he was estimating. If, unable to estimate the feelings of the other person, he merely used his own feelings as a guide, his accuracy score would have been high but not a true measure of social sensitivity.

An additional limitation of the social sensitivity questionnaire was its length. Sixteen questions are far too few, either psychologically or statistically, to obtain a truly valid sample of feelings or estimations. However, this limitation was imposed to keep the already difficult task of estimation from becoming so unpleasant as to threaten rapport. It was learned from a pilot study that when subjects found the estimation task too difficult, they stopped trying and marked their responses in random fashion. It was for this reason that the number of items on the questionnaire was kept small.

DISCUSSION

Accuracy as a Measure of Social Perception

The accuracy score as it was used in this study (a numerical representation of the proximity of an estimation to a criterion supplied by the estimatee) is apparently an unrefined index of social perception, though not wholly without reliability. Significant differences in accuracy among the items of the questionnaire testify to their lack of homogeneity. There was, for example, a substantial difference in accuracy between the negatively and positively phrased items, which was checked by the chi-square test ($p < .07$). Presumably there were many other differences that contributed to the lack of homogeneity among the items. Had these differences been consistent among estimators, estimatees, or groups, there would be some basis for evaluating an item according to its ease of estimation. But this was not the case. Items that were difficult to estimate in one group were easy in another. Estimators varied considerably as to which items yielded them the greatest accuracy, and estimatees varied in their relative estimatability on the different items.

A given item does not seem to have a fixed level of difficulty. Rather, it depends upon the situation in which it is applied. The job of estimating whether a person felt nervous would vary in difficulty, depending on how nervous

he actually appeared to be. If he bit his nails or trembled, the particular item would be easy for an estimator. The same holds true for groups as a whole. If a group verbally patted itself on the back for the job it had done, group accuracy on Item 2, "I was satisfied with what we accomplished," would be high. The variability of the estimation task itself offers an explanation for the significant group and individual differences.

In a sense, this kind of estimation task is very realistic. In everyday interaction, social perception is influenced not only by the personal characteristics of the individuals involved, but by the content of the social situation itself. However, a more analytical approach to the accuracy score and its components requires a better controlled set of criteria that would provide a standard basis for comparing estimations.

Cronbach (2) and Gage (4) are at present attempting to break down the accuracy score into several meaningful components, while Bronfenbrenner and Harding (1) have hypothesized the existence of at least two distinct and independent types of social perception: one, the ability to predict the normative responses of a given population, and the other the ability to predict a differential response among individuals. The former has already been demonstrated by Gage to be a surprisingly important factor in predictive accuracy.

Sources of Variables Affecting Accuracy

Social sensitivity in small interactive groups seems to be a function of three kinds of variables: personal variables associated with the characteristics of the individual, interpersonal variables associated with the relationship between individuals, and group-situation variables associated with the group as a whole and its social and physical characteristics.

Personal variables. The highly significant accuracy differences among estimators and estimatees demonstrate that personal variables influence a person's predisposition to be accurate and estimatable. The traditional monadic concept of sensitivity must be expanded to include the very significant role of the estimatee whose traits are as much responsible for accurate estimations as those of the estimator. Furthermore, there is a substantial correlation between accuracy and estimatability for a given person, which suggests the possibility that these functions have a common or at least related source. Perhaps this is analogous to the practice of telegraphy. Although sending and receiving Morse code are separate and distinct operations, nobody would question the fact that these two skills are highly related and do in fact have so many common determinants as to be almost two aspects of the same skill.

Attempts to correlate social perception with particular personal variables have been numerous. The results have been inconclusive largely because experimenters used different measures of social perception. In the present study it was hypothesized that two characteristics of an individual would influence his ability to make accurate estimations: (*a*) his feelings toward himself, and (*b*) his feelings toward others. The results did not substantiate the first, but did support the second. The people who showed more favorability toward others were more accurate in estimating their feelings.

One explanation of these results hinges upon the findings of Fiedler (3) who has demonstrated that people assume greater similarity between themselves and persons they like than persons they don't like. The person who is more favorable toward others probably sees them as quite similar to himself. Therefore, if an estimator likes his associates, he assumes with a fair amount of accuracy that they in turn like him.

This raises the question as to whether the assumed similarity of an estimator is a distortion of social perception. If a group member has feelings that are quite typical of the group, his estimations based on assumed similarity would be extremely accurate. Some people have dismissed accuracy achieved in this fashion as artificial, not properly belonging to the realm of social perception. But where does one draw the line between assumed and recognized similarity? If it is true that an important component of accuracy is the ability to make correct judgments about the "group stereotype" (4) or "generalized other" (1), it would certainly make a given estimator more insightful if he were typical of the group to which his estimatee belonged.

Interpersonal variables. The results of the experiment indicate that when the self-criterion and other-criterion items are regarded sepa-

rately, a distinct interaction effect appears among the *S*s. That is, each *S* seemed to fare better when estimating the feelings of some people than others. This had no bearing on how successfully the others could estimate his feelings. Going back to the telegrapher analogy, it is undoubtedly the case that whether or not A is a good telegrapher, his speed and accuracy of receiving will depend somewhat upon the particular style of B, the sender. But this does not necessarily mean that if A receives well from B, B is going to find A's style easier to receive. The one has nothing to do with the other except that if A is a good sender, in general, one would expect him to be a good receiver as well. Accuracy, then, is apparently a function of the relationship between the sender and the receiver of feelings and the degree to which either or both possess a certain skill or potentiality in the communication of these feelings.

If an estimator's favorability toward others is correlated positively with his perception of their feelings toward him, one would expect that a pair of people who were favorably oriented toward each other would also be more highly perceptive of each other. This was the hypothesis with respect to reciprocal favorability and accuracy within pairs. The correlation (.54) found in the data supported the hypothesis but was not significant enough to be conclusive. Further investigation with a larger sample is indicated, since the same kind of relationship was found between the two variables at the individual and group levels.

Out of the two measures of similarity within pairs, the only one that correlated significantly with the combined accuracy of the pair was the measure based on their criterion ratings of each other. Here is another case where people who had similar favorable feelings toward each other were highly perceptive of these feelings. But why should not people who respond to each other similarly, regardless of whether or not the response is favorable, recognize each other's feelings? The answer to this question may lie in a concept expressed by Sears (6). "... anticipatory reactions ... are the *expectancies* that make the behavior of two people truly interdependent. They provide the mechanism, at least within the framework of one conceptualization of the learning process, by which a diadic behavior unit can be derived from the combining of two or more monadic units." To transpose this concept to a concrete situation, imagine two strangers suddenly placed together, similar in thoughts and feelings but lacking, so far, any kind of diadic or interactive relationship. At this point their social perceptions of each other are inaccurate. Sears suggests that they have not developed a set of expectancies toward each other. Without mutual expectancies the pair do not recognize how they are similar or that they are similar at all. They have no "real" basis upon which to make an estimation of each other's feelings (outside of their knowledge of people in general). They have the potentiality for estimating each other's feelings with extreme accuracy, but until they have come to recognize this through the learning that accompanies interaction, and have developed reciprocal expectancies, they will be less accurate. Perhaps reciprocal favorability is a symptom of a more highly developed interpersonal relationship that provides the expectancies needed for high-level accuracy.

Group-situational variables. The fact that the three experimental groups differed significantly with respect to their total amount of intragroup accuracy does not answer any questions as much as it raises a new one: What might be responsible for these differences?

One explanation may be that the actual events of each group's interaction session were different, since the groups were given complete freedom as to how they were to plan their skit. As it was pointed out in the discussion of the items, it is likely that the events of one group's interaction session would bring feelings more out into the open and cause them to be more easily estimated than another group's; hence, the significant group-item interaction effect.

With respect to the correlates of accuracy at the group level, it is interesting to note that all the hypothesized relationships between sentiment structure and accuracy were supported by high and positive correlations. These results, though not conclusive, suggest that at the group level, expectancies are more easily formed. It would seem that after only 30 minutes of interaction, generalized group estimations would be more accurate than interpersonal perceptions. Although Gage (4) has demonstrated that people can draw upon well-established group expectancies to make fairly

accurate predictions of the responses of a member of the group, when the estimatee deviates from the group norm, predictions of his responses are likely to be inaccurate. But in estimating an entire group, the effect of these deviations is reduced and total accuracy depends more upon how well the norms of the group's behavior have been learned by its members.

SUMMARY

This study was undertaken to measure and analyze social sensitivity in the small, task-oriented group. An analysis of variance design was used to investigate the determinants of accuracy in three five-man groups. An analysis was made of the covariance between accuracy and the self-favorability, other-favorability, and similarity of sentiment structure among the *S*s. The following findings were reported:

1. Estimation accuracy is a function of (*a*) characteristics of the estimator, (*b*) characteristics of the estimatee, (*c*) the nature of the group and the situation, and (*d*) the particular feelings being estimated.

2. A person whose feelings toward the other members of his group are favorable tends to be more accurate in estimating their feelings.

3. A pair of group members whose feelings toward each other are favorable or similar, tend to be more accurate in estimating each other's feelings.

4. A pair of group members whose feelings toward themselves are similar, tend to be more accurate in estimating each other's feelings.

5. A group whose feelings toward one another are favorable or similar tend to be more accurate in estimating one anothers feelings.

6. A group whose feelings toward themselves are favorable or similar tend to be more accurate in estimating each other's feelings.

REFERENCES

1. BRONFENBRENNER, U., & HARDING, J. Components and correlates of social sensitivity. Paper read at Amer. Psychol. Ass., New York, September, 1954.
2. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
3. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381–388.
4. GAGE, N. L. Accuracy of social perception and effectiveness in interpersonal relationships. *J. Pers.*, 1953, **22** (1), 128–141.
5. HENDERSON, C. R. *General, specific, and maternal combining abilities among inbred lines of swine.* Unpublished doctor's dissertation, Iowa State Coll., 1952.
6. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist*, 1951, **6**, 476–483.

Received January 19, 1955.

THE MEASUREMENT OF HOSTILITY IN CLINICAL SITUATIONS[1]

ARNOLD H. BUSS, ANN DURKEE AND MARC B. BAER

Carter Memorial Hospital, Indianapolis

THIS study is part of a program of research that seeks to provide operational definitions for clinically important variables. The necessity for such research is illustrated by Grayson and Tolman's semantic study of clinical concepts. They had clinical psychologists and psychiatrists define 50 terms frequently used in psychological reports, e.g., *aggression, anxiety, dependent, hostility, rigidity.* The results indicated marked variability in definition, and the authors concluded:

> The most striking finding of the study is the looseness and ambiguity of the definitions of many of these terms. Up to the present time, verbalization and conceptualization have clearly lagged behind the useful clinical application of psychological terms (2, p. 229).

The ambiguity of most clinical concepts stems from the multiplicity of their behavioral referents. A previous study (1) examined the term *anxiety* and found that it referred to six different categories of behavior. Similarly, the term *hostility* appears to have several different meanings. In the Grayson and Tolman study it was defined variously as: (*a*) negative feeling states, (*b*) destructive impulses, (*c*) aggressive behavior, and (*d*) a reaction to frustration. Thus, the concept of hostility seems to consist of several subconcepts.

The present study seeks to identify and define the subconcepts of hostility and to investigate their use in measuring hostility in psychiatric patients. If the various aspects of hostility could be reliably evaluated in a clinical situation, such evaluations might be related to a more efficient, time-saving instrument. Accordingly, the Iowa Hostility Inventory was administered, and a secondary purpose of the study is to compare this inventory with clinical evaluations of hostility.

METHOD

Subjects

The *S*s were patients in a neuropsychiatric hospital, 30 men and 30 women. This was a diverse sample of a patient population, and there were many different diagnoses. Mental defectives and patients with known organic brain damage were excluded from the study, as well as patients who were so far out of contact as to be uncommunicative.

Procedure

Each patient was interviewed in a room with three psychologists, who served as judges. Every patient was interviewed by the same psychologist, although the other two psychologists also asked questions at the end of the interview. The interview lasted approximately 15 minutes. Then the Iowa Hostility Inventory was administered by another staff member.[2] Immediately after the interview each judge rated the patient for hostility, each rating being made independently and without consultation with the other judges.

The Interview

The interview followed a relatively uniform course. Initially, the patient was encouraged to discuss problems of adjustment at home and at work. Then the interviewer led the discussion into the areas of resentment, frustration, and distrust, and from there to the more behavioral aspects of hostility. At the end the questions tended to be more specific, e.g., "Do you ever slam doors when you are angry?" or "Do you sometimes curse?"

The interviewer attempted to maintain a warm, friendly attitude with all patients, and it is believed that he was successful, within limits. Some patients had to be encouraged to verbalize more than others, but it is felt that the variability in the interviewer's approach was not excessive.

The Hostility Ratings

The dimensions of hostility used in this study were derived from a rational analysis of hostile behavior. In everyday usage clinical psychologists differentiate between the various manifestations of hostility, and the categories to be described are the outcome of an attempt to rationalize this everyday clinical usage. Each area of hostile behavior was rated on a seven-point scale.

There were seven different categories of hostile behavior, as follows:

1. *Resentment.* This refers to a feeling of anger over real or fantasied mistreatment. This aspect of hostility tends to be covert, and it is usually verbalized in terms of "feeling" verbs, e.g., "I felt irritable," "I was annoyed at him." It is also reflected in whining, complaining, or demanding behavior.

2. *Verbal Hostility.* This category refers to the style of the patient's verbalizing rather than the content. Behaviors included are yelling, cursing, screaming, arguing, shouting, and being caustic, ironic, or teasing.

[1] This study was carried out with the aid and encouragement of Juul Nielsen, M.D., Superintendent. Thanks are also due to Dr. Ruth Hines and Dr. Edward Strain for their helpful criticism of the manuscript.

[2] We wish to thank Miss Beulah Gore for administering and scoring the inventory.

The emphasis here is on *how* the patient verbally interacts with others rather than on *what* he says.

3. *Indirect Hostility.* This refers to the expression of hostility in a roundabout or passive fashion. Included are negativism, recalcitrance, spiteful silence, slamming doors, and violent acts that affect the object of the hostility indirectly, e.g., smashing the hated person's car.

4. *Assault.* This category refers to overt acts of violence that are directed toward an object. Included are fighting with and striking others, breaking or destroying objects and defending against attack. Defense of self was included because some patients cannot manifest enough aggression to fight off an attacker.

5. *Suspicion.* This refers to an attitude of distrust of others in which hostility is projected onto other people. Included are beliefs that the patient is being "railroaded," persecuted, or punished, and verbalizations such as, "You can't trust anyone."

It should be noted that the first five categories may be classified into two groups, one contentual and the other stylistic. Resentment and Suspicion both refer to feelings, attitudes, or specific complaints; they deal with *what* the patient is hostile about rather than the manner in which he expresses hostility. On the other hand Verbal Hostility, Indirect Hostility, and Assault refer to the style or mode of the patient's aggressive behavior. If a given patient were resentful or suspicious, he might express it verbally, indirectly, assaultively, or not in any of these ways. It might be expected that the presence of Verbal Hostility would always be accompanied by a feeling of resentment. This is undoubtedly true in many cases, but it does not necessarily follow. It is possible for a patient to be extremely argumentative without necessarily having feelings of irritation and annoyance. In fact, in the present sample there were patients who tended to become involved in disputes, yet verbalized little or no resentment.

6. *Over-all Evaluation of Hostility.* This represented a pooling of the first five categories. The judges were instructed to attempt a subjective estimate of global hostility rather than merely to average the ratings of the individual categories.

7. *Strength of Hostile Urges.* In this study several clinically important aspects of hostility were omitted. No allowance was made for the mechanisms of reaction formation, avoidance, and somatization. These mechanisms are often used by patients in their attempts to deal with hostile urges. However, there is no general agreement on which somatic complaints are indicative of hostility, and it seems certain that variables other than hostility can account for many of the somatic symptoms usually associated with hostility. Similarly, reaction formation to aggressive urges is difficult to evaluate. The typical behaviors expected are oversolicitude, excessive concern, and a saccharine sweetness. Unfortunately, there is no clear-cut basis for differentiating between the concern of a good samaritan and the solicitude of a hostile person with a reaction-formation mechanism.

In light of these considerations an "intuitive" category was added, Strength of Hostile Urges. This category involved a rating of "hostile drive strength," whether or not it was expressed overtly. A patient who manifested great hostility would receive a high rating on Strength of Hostile Urges because such urges are obviously present. However, a patient who manifested little hostility might also be rated high on Strength of Hostile Urges on the basis of the judges' "clinical impressions." It was expected that the interrater reliability would be lower than that of the other categories because of the lack of specific denotation of the behavior involved. The Strength of Hostile Urges was not used in arriving at the over-all rating of hostility.

TABLE 1

INTERRATER CORRELATIONS AND CORRELATIONS BETWEEN EACH CATEGORY AND THE OVER-ALL RATING OF HOSTILITY*

Category	Interrater Correlations		Over-all	
	Mean	Range	Men	Women
Resentment	.77	.74–.81	.77	.84
Verbal Hostility	.85	.83–.87	.77	.87
Indirect Hostility	.88	.80–.95	.57	.68
Assault	.95	.95–.95	.75	.73
Suspicion	.88	.82–.95	.54	.78
Over-all	.84	.79–.89		
Strength of Hostile Urges	.57	.54–.64	.77	.84

* All correlations in this study are Pearson product-moment *r*'s. They were averaged by the *z* transformation method. All correlations in this table are significantly different from zero at the .01 level of confidence.

RESULTS AND DISCUSSION

In order to establish the reliability of the ratings, each judge's ratings were compared with those of the other two judges. The correlations are shown in Table 1. The mean intercorrelations reveal that the various aspects of hostility were assessed with a moderate to high degree of reliability, with a single exception. The exception was Strength of Hostile Urges, with a mean intercorrelation of .57. The rating in this category was intuitive, the judges being asked to give their "clinical impressions." There were no specific behaviors to assess, but only minimal cues that were often implicit and unverbalized. Since there were no specific, overt behaviors to assess, the comparative unreliability in rating Strength of Hostile Urges is not surprising. The remaining categories were reliably assessed, which suggests that the definitions are adequate.

Next, interjudge correlations were computed for men and women separately. It was found that the ratings for men were as reliable as the ratings for women, the correlations differing no more than a few points in each instance.

The best estimate of a patient's rating for any given category is the consensus of the three judges. Therefore, ratings of the three judges were pooled, and these pooled ratings

TABLE 2

CORRELATIONS BETWEEN RATINGS OF VARIOUS ASPECTS OF HOSTILITY AND THE IOWA HOSTILITY INVENTORY*

Category	Iowa Hostility Inventory	
	Men	Women
1. Resentment	.30	.51
2. Verbal Hostility	.26	.46
3. Indirect Hostility	.24	.29
4. Assaultiveness	.36	.41
5. Suspicion	.24	.56
6. Over-all	.39	.54
7. Strength of Hostile Urges	.37	.32

* For an N of 28 $p = .05$ when $r = .37$, and $p = .01$ when $r = .48$.

were used in all subsequent comparisons. Further, since sex differences might show up, e.g., men are considered to be more assaultive than women, all subsequent analyses were done for men and women separately.

In this study hostility was divided into six categories. Table 1 presents data on the relationship between each category and the over-all evaluation of hostility. For both men and women all the correlations between the over-all rating and the individual categories are significantly different from zero at the .01 level of confidence. The significant positive relationships suggest that all the individual aspects of hostility are important in making a global evaluation of hostility. In men the Over-all *vs.* Indirect and Over-all *vs.* Suspicion correlations are somewhat lower than the correlations between the other categories and Over-all. The Indirect and Suspicion categories represent the more subtle and covert aspects of hostility. Evidently, in men these aspects are not as important in the evaluation of global hostility as the more overt manifestations of hostility. On the other hand the women's correlations between Over-all and the individual categories do not differ greatly in magnitude. Apparently, in women the covert aspects of hostility are just as important as the more obvious aspects.

Next, the pooled ratings in each category were compared with the scores on the Iowa Hostility Inventory[3]; the correlations are presented in Table 2. In general, the women's correlations are higher than the men's. Only two out of seven men's correlations are significantly different from zero, while five out of seven women's correlations are significant. The correlations between Over-all Hostility and the Iowa Hostility Inventory are significant for both sexes, suggesting that hostility as measured in an interview is definitely related to hostility as measured in an inventory. However, the magnitude of the correlations indicates that the inventory cannot be substituted for the more cumbersome clinical appraisal of hostility.

SUMMARY

This study attempted to formulate operational definitions for the various meanings of the concept of hostility. Thirty men and thirty women, patients in a neuropsychiatric hospital, were interviewed and rated by three judges for the following aspects of hostility: Resentment, Verbal Hostility, Indirect Hostility, Assault, Suspicion, Over-all Hostility, and Strength of Hostile Urges. Also, the Iowa Hostility Inventory was administered.

It was found that the judges' ratings of the various aspects of hostility showed general agreement. All the individual aspects were related to Over-all Hostility. The correlation between Over-all Hostility (interview) and the Iowa Hostility Inventory was .39 for men and .54 for women, suggesting that the Inventory is not a substitute for a clinical evaluation of hostility.

REFERENCES

1. BUSS, A. H., WIENER, M., DURKEE, ANN & BAER, M. The measurement of anxiety in clinical situations. *J. consult. Psychol.*, 1955, **19**, 125–129.
2. GRAYSON, H. M., & TOLMAN, RUTH S. A semantic study of concepts of clinical psychologists and psychiatrists. *J. abnorm. soc. Psychol.*, 1950, **45**, 216–231.

Received January 21, 1955.

[3] Two men and two women failed to complete the inventory, leaving N's of 28 for men and 28 for women.

PERCEPTUAL CONTACT WITH REALITY IN SCHIZOPHRENIA[1, 2]

EDWARD LOVINGER

The RAND Corporation, Santa Monica, California

In most textbook or theoretical discussions of schizophrenia one encounters either the implication or a direct statement that schizophrenics are in poor contact with reality (1, 5, 6, 7, 9, 11). Generally, a precise definition of contact with reality is not given, but frequently in many of these discussions there is the implication that the schizophrenic's break with reality includes basic perceptual processes, i.e., such phenomena as perception of size, shape, color, pressure, pain, etc. Much of the work with projective techniques certainly suggests this (2, 3, 4, 10, 20). However, since these tasks usually reflect more complex psychological functions, it is difficult to ascertain to what extent basic perceptual processes are involved.

For a number of reasons, it was felt that a simple size-constancy experiment would be a very effective means of testing this implication: (*a*) it involves a relatively simple task that even schizophrenics in poor contact with reality can perform, (*b*) size constancy represents a basic and universal perceptual process, (*c*) it lends itself to rigorous experimental control and quantification, and (*d*) it has been clearly shown that as distance cues are reduced, size constancy is also reduced (12, 13, 15).

The last point is important in that it would seem to be a reasonable assumption that if an individual is in poor contact with reality in terms of basic perceptual processes, he then would be less responsive to minimal distance cues as compared to an individual in good contact with reality. Granting this assumption and considering the fact that size constancy decreases as distance cues are reduced, one would anticipate that when distance cues are minimal, these individuals would manifest less size constancy than individuals in good contact. Hence, if schizophrenics described as being in poor contact with reality are *perceptually* in poor contact, one should be able to predict that under experimental conditions involving minimal distance cues, they will manifest less size constancy then either normals or schizophrenics in good contact.

Two investigations of size constancy in schizophrenia have been reported in the literature. Neither Raush (18) nor Sanders and Pacht (19) found any reduction in size constancy for schizophrenics. This is not surprising in that in both investigations the schizophrenic subjects employed were in relatively good contact with reality. Moreover, the experimental conditions were such that the distance cues available were much more than minimal. These factors were taken into account in the design of the present experiment.

It was the purpose of the present investigation to determine whether schizophrenics considered clinically to be in poor contact with reality are *perceptually* in poor contact with reality as shown by their performance in a simple size-constancy experiment. It was predicted that under experimental conditions involving minimal distance cues, schizophrenics considered in poor contact with reality would manifest less size constancy than either normals or schizophrenics considered in good contact.

METHOD

Apparatus. The apparatus consisted primarily of two light-tight tunnels of fixed but different lengths, one 13 ft. and the other 4 ft. The height and width of each tunnel was 1.5 ft. The tunnels were mounted on stands which were 41 in. in height. Within each tunnel, at the end farthest from the subject, there was a 1.5-ft. square sheet of flashed-opal plate glass placed 1 ft. from the far end. The target-light source was a 60-w. incandescent bulb behind each glass plate. Within each tunnel, 9.5 in. from the front end, there were two 60-w. incandescent bulbs, one on each side, which could serve to illuminate the tunnels. These lights were hidden from the view of the subject by means of black shields which were semicylindrical in shape and ex-

[1] From the Veterans Administration Hospital, Battle Creek, Michigan.

[2] This article is based upon a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Michigan State College. The writers wishes to acknowledge his indebtedness to Professors G. M. Gilbert, M. R. Denny, and D. M. Johnson. The writer is particularly indebted to Dr. Paul Greenberg for his invaluable aid in the preparation of this article. Detailed information concerning the apparatus employed and the data obtained may be found in the original thesis (16).

tended from the top to the bottom of the tunnel. They did not appreciably restrict the *S*'s view of the tunnels.

The targets were uniformly lighted, clearly defined, disks of light at the end of each tunnel. This was achieved by inserting strips of sheet metal (called target plates) with varying sized, precision-drilled circular holes. These target plates were painted flat black to conform with the interior of the tunnels and except for the circular hole completely covered the flashed-opal glass. The target plates were inserted in the top of the tunnels directly in front of the flashed-opal glass; thus, the target in the long tunnel was 12 ft. from the front end, and 3 ft. from the front end in in the short tunnel. The circular hole in each of the target plates was so located that, when the target plate was inserted, the disk of light it defined was centered on the flashed-opal glass. There was only one *standard* target plate, which produced a standard target 2 in. in diameter. There were 24 *comparison* target plates. The holes in these ranged from $\frac{1}{8}$ in. to 3 in. in diameter, in $\frac{1}{8}$-in. step intervals. The comparison target plates were numbered consecutively from 1 to 24 in order from smallest to largest. Thus, the number of each comparison target plate represented, in eighths of an inch, the diameter of the target it defined (target plate 16, e.g., defined a target $\frac{16}{8}$ or 2 in. in diameter). The standard target was always in the long tunnel and the comparison in the short tunnel.

When the target plates were being changed, all apparatus lights were turned off in order to control cues between trials. A variable transformer controlled the illumination within the tunnels and thus, indirectly, the distance cues. A rheostat regulated the level of illumination of the target lights, and thus the brightness of the targets.

There was a felt-lined, light-tight slot in which could be inserted a reduction screen 9 in. from the front end of the long tunnel. The reduction screen had an aperture $\frac{3}{16}$ in. in diameter. With the reduction screen in place, only the target, or little more, was visible to the subject, i.e., he could not see the tunnel itself.

At the front of each tunnel there was a stereoscope viewer that could be adjusted for binocular or monocular vision. These viewers were padded with foam rubber in order to make the tunnels light tight. Only the stereoscope viewers could be seen by *S* since the rest of the apparatus was hidden from his view by an opaque cloth screen.

Subjects. There were three groups of subjects (*S*s): 18 male schizophrenic patients considered clinically in good contact with reality (for the most part these patients were in partial remission), 18 male schizophrenic patients considered in poor contact with reality, and 18 normal controls. The two schizophrenic groups were patients at the Battle Creek Veterans Administration Hospital. The controls were patients carrying a physical diagnosis at the Saginaw General Medical and Surgical Veterans Administration Hospital. The three groups were equated for age, ranging from 22 to 42, with a mean of 32. All *S*s had been tested for visual acuity and had at least 20/25 vision, corrected or uncorrected. Those *S*s whose vision had been corrected were required to wear their glasses throughout the test procedure. Each of the schizophrenic groups consisted of nine patients diagnosed schizophrenia, paranoid type, and nine patients diagnosed schizophrenia, but a type other than paranoid, i.e., each group contained nine paranoid schizophrenics and nine nonparanoid schizophrenics.

The selection criteria for the good-contact schizophrenics were: (*a*) unanimous agreement among three judges (the experimenter, the ward charge attendant, and the patient's physician) that the patient was in *good contact* with reality; (*b*) open-ward assignment; (*c*) diagnosis of schizophrenia and the particular subclassification by the hospital psychiatric staff; (*d*) diagnosis consistent with case history material and available psychological test material; (*e*) no implications of mental deficiency or neurological involvement; (*f*) neither insulin nor electroshock therapy within the preceding three months. The selection criteria for the poor contact schizophrenics were: (*a*) unanimous agreement among three judges (the experimenter, the ward charge attendant, and the patient's physician) that the patient was in *poor contact* with reality; (*b*) closed-ward assignment; (*c–f*) same as for the good contact group. The selection criteria for the normal group were: (*a*) no history of emotional difficulties which necessitated either hospitalization or psychiatric consultation; (*b*) no indications of mental deficiency or neurological involvement; (*c*) no diagnosis of psychosomatic disorder or any illness which might have a psychological basis.

Procedure. Each *S* was tested under three different cue conditions which shall be referred to as the maximal cue, minimal cue, and no-cue conditions:

A. Maximal cue—the targets were presented with both tunnels fully lighted while the subject viewed them binocularly. The brightness of each of the tunnel lights, i.e., those at the front of the tunnel rather than the target lights behind the flashed-opal glass, was 1200 ft.-l. as measured by a type DW-68 General Electric light meter 1 ft. from the source.

B. Minimal cue—the targets were presented with both tunnels very dimly illuminated while the *S* viewed them monocularly. The light intensity of each of the tunnel lights under this cue condition was 3 ft.-c. with the light meter 5.5 in. from the light source.

C. No-cue—the targets were presented with the tunnel lights off while *S* viewed the targets monocularly. Throughout this condition the reduction screen was in place in the longer tunnel. The only light in the tunnels was that from the targets.

The light intensity of the target-light sources was maintained at a very low level throughout the experiment in order to minimize the amount of stray light from them. Their light intensity was 1 ft.-c. with the light meter 1 in. from the light source.

Each *S* was tested on three successive days with one cue condition each day. The order of presentation of the three cue conditions was systematically randomized. Thus, all groups received all possible orders of cue conditions.

The *S* was brought into the testing room by an assistant who remained at the front of the apparatus with him throughout the testing. The *E* and an assistant were stationed at the rear of the apparatus where they could change the comparison target plates, insert or remove the reduction screen, or change the level of illumination within the tunnels. The *S* was seated on an

adjustable stool so that his eyes were level with the viewers. For all *S*s the standard was in the long tunnel and the comparison in the short. The directions given *S* stressed a naive, phenomenal attitude. He was instructed by the assistant at the front of the apparatus to tell whether the comparison target *looked* larger, smaller, or the same size as the standard. In every instance, *S* was required to keep his face tightly against the viewer when looking at the target. This procedure was followed in order to insure that no stray light entered the tunnels via the viewers and that *S* obtained no additional cues from head movements. When an *S* indicated that the two targets were the same size, the number of the comparison target plate was recorded as his choice and this was considered as a trial. Thus, for each trial, the diameter of the comparison target *S* chose as a match for the standard was recorded in eighths of an inch.

Every *S* was given five trials for each cue condition. The first trial for each cue condition was considered a practice trial and was not included in the experimental data. The presentation of the targets was in order of their sizes, i.e., from smallest to largest or largest to smallest. For the first trial, the targets were presented in ascending order of size. For subsequent trials, the order of presentation was alternately descending or ascending. In order to increase the probability that *S* was responding to the targets themselves rather than the ordinal position of the targets, the initial targets for the trials differed, according to a predetermined plan (16), within and between cue conditions.

RESULTS

The data were first subjected to an analysis of variance. The technique used was essentially that described by Edwards (8, p. 295) for analysis of data involving repeated measurements on the same *S*s. The results of this analysis of variance did not indicate any significant differences among the groups in their mean constancy judgments for any of the cue conditions. However, the *F* for the orders of presentation of the cue conditions was significant beyond the .001 level of confidence. Order not only affected the degree of constancy, but, also, increased the variance within groups. Moreover, the Groups × Orders × Cue conditions interaction was significant beyond the .01 level of confidence, indicating that these variables were interacting in a very complex fashion.[3] It was felt that this complex interaction plus the increased variance within groups due to the order effect might be masking any differences that existed between the groups.

On this basis, a second analysis of variance was made with the order effect taken out. For this analysis, only the data from the first cue condition that each *S* received were utilized. Hence, the order of presentation of cue conditions could have no effect upon these data. The technique employed was essentially that described in Edwards (8, p. 208) for analysis of data involving factorial designs. The *F* for the differences between the groups for the three cue conditions combined was significant ($p < .05$). The Groups × Cue conditions interaction term was not significant, indicating that the differences between the groups did not vary significantly among the three cue conditions. This finding, when considered with the fact that there were significant differences among the groups for the three cue conditions combined, suggests that there were significant differences between the groups on the three conditions taken separately.

It may be seen by inspecting Tables 1 and 2 that with the order effect removed there were no significant differences between the means of the groups for either the maximal cue or no-cue conditions. In regard to the minimal cue condition, the mean of the poor contact group was

[3] No consistent nor meaningful relation with regard to the groups was found in this interaction.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF COMPARISON TARGETS SELECTED

(diameter in eighths of an inch)*

Group	Maximal Cue Condition		Minimal Cue Condition		No-cue Condition	
	M	*SD*	M	*SD*	M	*SD*
Good contact	16.96	2.14	7.21	.71	4.54	.60
Poor contact	15.38	1.74	5.54	.98	4.58	.82
Normal	16.29	1.70	6.75	.65	4.75	.45

* Four equivalent to a retinal match and hence, to zero size constancy; 16 equivalent to an objective match and hence to complete size constancy.

TABLE 2

t RATIOS OF GROUP DIFFERENCES

Group	Maximal Cue Condition	Minimal Cue Condition	No-cue Condition
Good contact vs. poor contact	1.41	3.41*	.04
Good contact vs. normal	.60	1.18	1.00
Poor contact vs. normal	.92	2.52**	.18

* $p < .01$ with one-tailed *t* test.
** $p < .02$ with one-tailed *t* test.

significantly lower than either the good contact or normal groups.[4] Thus, the prediction that under experimental conditions involving minimal distance cues, schizophrenics considered clinically in poor contact with reality will manifest less size constancy than either schizophrenics in good contact or normals was supported by the data.

The data were also analyzed for differences in constancy between the paranoid and nonparanoid schizophrenics. Analyses of variance both with the order effect in and removed did not reveal any significant differences between these subgroups, either for the cue conditions combined or considered separately.

DISCUSSION

When the order effect was removed, it was found that under the minimal cue condition schizophrenics considered in poor contact with reality manifested significantly less size constancy than either normals or schizophrenics in good contact. This finding strongly suggests that the schizophrenics break with reality involves not only more complex psychological functions, but basic perceptual processes as well. Moreover, this finding highlights a point often made in discussions of schizophrenia, but frequently ignored in research involving schizophrenics, namely, one cannot treat schizophrenia as an homogeneous entity.

Should these findings be substantiated by further research, they would appear to have considerable theoretical and practical importance. One implication is that they lend some support to psychoanalytic theory. Thus, Bruner (7) suggested that the characterization of schizophrenia by psychoanalytic theory as a regression to primary narcissism could be tested by means of a perceptual experiment. He proposed that one might predict from psychoanalytic theory a "breakdown in such phenomena as size and shape constancy" (7, p. 142). Tentatively, this investigation appears to give some support for this view of psychoanalytic theory. However, it should be noted that these results lend themselves readily to another interpretation. Hunt and Cofer (14) and others have pointed out that much of the data supposedly indicating regression in schizophrenia could be accounted for on the basis of motivation, attention, cooperativeness, etc. These variables, particularly attention, could account for these findings. A second implication of these findings would lie in the support which they lend to the view, as typified by Angyal, Freeman, and Hoskins (1), that withdrawal in schizophrenia is global rather than segmental and is manifested in a variety of psychological and physiological characteristics. The knowledge that the schizophrenic's poor contact with reality extends to basic perceptual processes could account, in part, for many of their bizarre responses on projective techniques, particularly on such tests as the Bender-Gestalt, Draw-a-Person, etc. Furthermore, much of the anecdotal material concerning the schizophrenic's misperceptions or lack of perception might be, in part, accounted for by this.

Although Raush (18) found that his paranoid schizophrenics obtained higher size-constancy scores than the nonparanoid schizophrenics, there were no significant differences between the paranoid and nonparanoid schizophrenics in this study. Raush felt that his nonparanoid group could be described as being more regressed and in poorer contact with reality than the paranoids. Thus, perhaps Raush's findings may have been more a function of contact with reality rather than of the diagnostic subgroups.

This study suggests the need for a number of investigations. If the schizophrenic's poor contact with reality extends to basic perceptual processes, one would anticipate a reduction in constancy for such phenomena as shape, color, and brightness constancy under experimental conditions involving minimal cues. Investigations of perceptual contact with reality in terms of other basic perceptual processes might prove fruitful.

SUMMARY

The purpose of this study was to investigate perceptual contact with reality in schizophrenia. It was assumed that if schizophrenics considered in poor contact with reality were perceptually in poor contact, they would be less responsive to minimal distance cues. Hence, it was hypothesized that under experimental conditions involving minimal distance cues, schizophrenics in poor contact would manifest

[4] The one-tailed *t* test was employed in these particular comparisons since the direction of the difference had been predicted in the hypothesis.

less size constancy than either schizophrenics in good contact or normals. Three groups of subjects—schizophrenics in good contact with reality, schizophrenics in poor contact, and normals—were tested in a size-constancy experiment under three different distance cue conditions—maximal, minimal, and no-cue.

The results of the investigation supported the hypothesis. This was interpreted as suggesting that the schizophrenic's break with reality involves not only more complex psychological functions, but basic perceptual processes as well. Some suggestions for further research were discussed.

REFERENCES

1. ANGYAL, A., FREEMAN, H., & HOSKINS, R. G. Physiologic aspects of schizophrenic withdrawal. In S. S. Tomkins (Ed.), *Contemporary psychopathology*. Cambridge: Harvard Univer. Press, 1943.
2. BECK, S. J. Personality structure in schizophrenia. *Nerv. ment. dis. Monographs*, 1938, No. 63.
3. BECK, S. J. Errors in perception and fantasy in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley, Los Angeles: Univer. California Press, 1944.
4. BELL, J. E. *Projective techniques*. New York: Longmans, Green, 1948.
5. BELLAK, L. *Dementia praecox*. New York: Grune & Statton, 1947.
6. BLEULER, E. *Dementia praecox or the group of schizophrenias*. New York: International Universities Press, 1950.
7. BRUNER, J. S. Personality dynamics and perceiving. In R. Blake and G. Ramsey (Eds), *Perception—an approach to personality*. New York: Ronald, 1951.
8. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
9. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
10. FRIEDMAN, H. Perceptual regression in schizophrenia: an hypothesis suggested by the use of the Rorschach Test. *J. genet. Psychol.*, 1952, **81**, 63–98.
11. GOLDSTEIN, K. The significance of psychological research in schizophrenia. In S. S. Tomkins (Ed.), *Contemporary psychopathology*. Cambridge: Harvard Univer. Press, 1943.
12. HOLADAY, R. E. Die GrössenKonstanz der Sehdinge bei Variation der inneren und äusseren Wahrnehmungs bedingungen. Part II in E. Brunswik (Ed.), Untersuchungen uber Währnehmungsgegenstände. *Arch. f. d. ges. Psychol.*, 1933, 88, 419–486.
13. HOLWAY, A. H., & BORING, E. G. Determination of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21–37.
14. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
15. LICHTEN, W., & LURIE, S. A new technique for the study of perceived size. *Amer. J. Psychol.*, 1950, **63**, 280–282.
16. LOVINGER, E. Perceptual contact with reality in schizophrenia. Unpublished doctor's dissertation, Michigan State College, 1954.
17. LOVINGER, E. Note on Sanders and Pacht's investigation of size constancy of clinical groups. *J. consult. Psychol.*, 1954, **18**, 226–228.
18. RAUSH, H. L. Perceptual constancy in schizophrenia. *J. Pers.*, 1952, **21**, 176–187.
19. SANDERS, R., & PACHT, A. R. Perceptual size constancy of known clinical groups. *J. consult. Psychol.*, 1952, **16**, 440–444.
20. SCHAFER, R. *The clinical application of psychological tests*. New York: International Universities Press, 1948.

Received January 21, 1955.

ADDITIONAL OBSERVATIONS ON THE SOCIAL PERCEPTIONS OF AUTHORITARIANS AND NONAUTHORITARIANS

ALVIN SCODEL
The Ohio State University

AND

MARIA LIVIA FREEDMAN[1]
The University of Washington

In a recent study Scodel and Mussen (4) demonstrated that nonauthoritarian individuals make more accurate judgments about authoritarians than authoritarians make about nonauthoritarians. Authoritarianism and nonauthoritarianism were defined by high and low scores on the California F scale (1). Specifically, when an authoritarian and a nonauthoritarian were paired in a social situation, subsequent judgments of nonauthoritarians by authoritarians resulted almost invariably in the imputation of authoritarian attitudes to the nonauthoritarians. Subsequent judgments by nonauthoritarians of authoritarians tended to be distributed over the accurate, inaccurate, and middle categories; some authoritarians were judged accurately (i.e., as authoritarian), others were judged inaccurately (nonauthoritarian), and about half of the authoritarians were judged to be neither authoritarian nor nonauthoritarian.

It has been suggested by Crutchfield (2) that the nonauthoritarians may be merely expressing a stereotype of their own that others are more authoritarian than themselves. If true, nonauthoritarians should, then, perceive other nonauthoritarians to be more authoritarian than they actually are. The present study was designed to test this hypothesis by extending the procedure of the Scodel-Mussen experiment to a pairing of nonauthoritarians with other nonauthoritarians. In addition, a further control on the findings of the previous study was introduced by pairing authoritarians with other authoritarians.

Method

The method here was almost identical with the one used in the Scodel–Mussen study. Authoritarian and nonauthoritarian samples were selected from a group of 237 students in the introductory psychology course at the Ohio State University who had responded to the California F scale. In addition to the 28 F items, 30 items from the Minnesota Multiphasic Personality Inventory (MMPI) were also included in the questionnaire. In a minor departure from the earlier study, these items were intended merely to serve as "buffer items" for the F-scale items. On the second administration of the F scale, a group of different MMPI items was included in an attempt to disguise the sameness of the F-scale items.

In the administration of the initial questionnaire, subjects (*S*s) were told that they were responding anonymously, but each could be identified by date of birth, religion, and major field, all of which were entered on the questionnaire.[2] The 40 *S*s receiving the highest F scores (the authoritarian or high group) and the 40 *S*s receiving the lowest F scores (the nonauthoritarian or low group) were selected for the experiment. The cut-off points for the high and low groups were 115 and 79, respectively. The cut-off point for the high group in the earlier study had been 119, but a somewhat smaller frequency of high F scores in the present population necessitated this minor downward adjustment in the selection of highs. The present study, as did the first, excluded minority group members, i.e., Jews, Negroes, and Orientals.

Because of scheduling difficulties, the actual sample was finally reduced to 24 lows and 24 highs. The high group consisted of 12 males and 12 females and the low group of 9 males and 15 females. Unlike the initial study in which pairs were matched for sex, exigencies of scheduling in the present study precluded the matching of sexes in the experimental situation, but data to be presented in the following section indicate that matched and unmatched groups do not differ appreciably. Among the highs, five pairs were matched and seven unmatched, and among the lows, seven pairs were matched and five unmatched.

In the experimental situation, which was a social one, each pair of high *S*s and each pair of low *S*s were instructed to carry on a 20-minute conversation on radio, television, and movies. The *S*s were told that they had been selected to participate in an experiment investigating the nature of social interaction between people meeting for the first time. The experimenter (*E*) remained in the room while the conversation was in progress and made every possible attempt to keep the conversation centered around the above topics.

After the time limit was up, each *S* was taken to a separate room and given the F scale (with the additional MMPI items) again. This time, however, *S* was asked to respond to the questionnaire as he felt the other *S* with whom he had just been speaking would respond to it. There were no indications of recognition of the similarity between the two questionnaires.

At the end of the experiment, then, each *S* had responded twice to the F scale. The first administration gave the *S*'s actual F score; the second administration involved a kind of role playing, and from it the estimated F score of the partner was obtained. Actual and

[1] Formerly of Ohio State University.

[2] It was felt that anonymity would produce more honest responses, particularly since the questionnaires were administered in a classroom situation.

estimated (attributed to the partner) F scale scores were obtained by adding a constant of 4 to the value which had been assigned to each F item, e.g., a −1 response to an item was scored 3, a +2 response was scored 6.

RESULTS

The differences in the high and low groups between pairs matched and unmatched for sex are presented in Table 1.

Despite the small *N*'s, the fact that the means for the two groups are almost identical argues against any appreciable effect of sex on estimated scores.

Table 2 gives the results of the present study when the estimated scores were classified as (*a*) accurate (estimated score was in the same category as the actual score); (*b*) middle (estimated score was between 80 and 114); and (*c*) inaccurate (estimated score was in the opposite category, i.e., a low's perception of another low gave the latter a total of 115 or above, and a high's perception of another high gave that person a total score of 79 or below).

The chi-square test applied to this table yields a chi-square value of 14.45, which is significant ($p < .01$) and indicates that, in general, highs tend to make accurate estimates of other highs whereas lows tend to make inaccurate estimates of other lows. The importance of this result lies in its apparent contradiction to the Scodel-Mussen study. There, it will be recalled, highs perceived lows inaccurately whereas lows perceived highs accurately. However, the high and low *S*s of both studies seem to behave consistently regardless of who the partner is that is being estimated. The estimates of the highs in the two studies are presented in Table 3.

The chi-square test of significance applied to this table yields a chi-square value of 7.69 which is significant ($p = .05$). However, an interpretation of a real difference between the groups is mitigated by inspection of the raw data of the present study which revealed that only one high estimated another high to be below 100. The remaining seven scores in the middle range are close to the cut-off point of 115 and, in fact, two are 114. (The means of scores attributed to partners in the two studies were almost identical—124.19, in the first, and 122.33, in the second, with sigmas of 11.08 and 12.13, respectively. A *t* test yielded a value of .57 ($p > .50$).

TABLE 1

ESTIMATED MEAN SCORES OF MATCHED AND UNMATCHED GROUPS

Group		Mean	Sigma	*t*	*p*
Highs	M*	120.6	9.17	.63	.50
	UM**	123.5	13.73		
Lows	M	98.8	17.41	.30	.70
	UM	101.1	18.84		

* Matched for sex.
** Unmatched for sex.

TABLE 2

NUMBER OF ACCURATE, MIDDLE, AND INACCURATE ESTIMATES OF F SCORES MADE BY HIGHS AND LOWS

Group	Accurate	Middle	Inaccurate
Highs	16	8	0
Lows	4	14	6

TABLE 3

NUMBER OF SUBJECTS PLACED IN THE HIGH, MIDDLE, AND LOW CATEGORIES BY HIGHS IN THE SCODEL-MUSSEN STUDY AND THE PRESENT STUDY

Group	High	Middle	Low
High-Low*	26	1	0
High-High**	16	8	0

* Highs' estimates of lows.
** Highs' estimates of highs.

TABLE 4

NUMBER OF SUBJECTS PLACED IN THE HIGH, MIDDLE, AND LOW CATEGORIES BY LOWS IN THE SCODEL-MUSSEN STUDY AND THE PRESENT STUDY

Group	High	Middle	Low
Low-High*	9	13	5
Low-Low**	6	14	4

* Lows' estimates of highs.
** Lows' estimates of lows.

The estimates of the lows in the two studies are presented in Table 4.[3] The obtained chi-

[3] The high and middle categories are not identical for either of the statistical comparisons in Tables 3 and 4 since, in the Scodel-Mussen study, the lower limit for the high group was 119, and in the present study, because of the difficulty in obtaining an adequate number of high subjects, it was 115. However, this difference does not seem large enough to preclude comparisons.

square value of .57 is obviously not significant. The table suggests that lows as a group are less uniform in their perceptions of others, but that there is a tendency to attribute middle scores to others whether these others are high or low.

DISCUSSION

The results from *S*s with high F scores seem unequivocal enough. Of a total of 51 highs in the two studies, not one estimates his partner to be in the low range and only nine estimates (most of which border on high estimates) are in the middle range. Accordingly, it would seem that the perceptions of highs are governed by a "same-stereotype." In the present study, these perceptions happen to be fairly accurate since each partner was, in fact, high, but this accuracy cannot be attributed to insightful perception. The present data are consistent with the assertion of the earlier study that "In a social situation with a member of a peer group, he maintains his security by assuming that the values and attitudes of others are similar to his own" (4, p. 181).

But what of the lows? It was also stated in the earlier study that "the nonauthoritarian personality, on the other hand, has less need [than the authoritarian personality] to perceive similarity between his own and others' values and attitudes. His reactions to others will be more highly individualized and his evaluations more objective. Therefore, he will be better equipped to use subtle personality cues and make more personalized, insightful assessments of the attitudes and values of others" (3, p. 181). One of the results of the earlier study (9 accurate, 13 middle, and 5 inaccurate judgments of highs by lows) was interpreted as partial support for this hypothesis. Certainly, the present data are in contradiction and support, rather, Crutchfield's suggestion that lows may be expressing a stereotype of their own. This "other-stereotype" is neither as extreme nor as frequent as the "same-stereotype" of the highs, but it does reflect a tendency on the part of lows to be guided by extrarational considerations. Of 24 lows in the present study, 20 were estimated as higher than they actually were. In discussing the "jungle-character" of the world as a feature of the highs' ethos, Frenkel-Brunswik states, "the opposite variant was defined as *trustingness and openness* (Italics author's), as manifested by seeing people as essentially 'good' until proved otherwise; it was expected to be found predominantly in the low scoring subjects" (1, p. 411). Interview material was interpreted as supporting this statement. It would seem, however, that the reluctance of the lows to perceive similarities between their attitudes and the attitudes of others is evidence of a disposition to view others as guilty until proven innocent.

One possible explanation of the lows' estimates is that many low-scoring *S*s take on mildly paranoid attitudes as a result of the contemporary *Zeitgeist*. Such lows may perceive themselves to be part of a distinct minority whose "liberalism" is unpopular and the object of considerable attack and derision. If so, all strangers with whom one has had only brief and casual contact would be viewed as potential threats until guarantees of the "rightness" of their attitudes were provided.

Another explanation is that many lows may consider themselves a minority, but one which is select or elite. A conviction of intellectual superiority may be accompanied by a belief that others are comparatively obtuse, and hence, incapable of assuming the same attitudes.

There is still another and, undoubtedly, a more innocuous explanation. Conceivably, many persons are reluctant to make judgments of others in the absence of detailed information, even though cues for the making of such judgments are available. Many lows may prefer to err on the side of caution, and the safest guess under such a circumstance is to place the other person at or near the mean of whatever is being judged. The lows' estimates, then, may not indicate criticality of others, but, rather, the adoption of a "wait and see" attitude.

A final possibility is that lows actually have no stereotypes of others at all, but believe, rather, that most people will differ from themselves in a variety of ways, including answers to the F scale. Because of their high tolerance for individual differences, they may trust others despite a perception of differences. This explanation cannot be refuted, but it does seem unlikely. In a situation involving what is ostensibly a peer relationship between two people with equal status and from a similar subculture, a perception of attitudinal differences

implies some sort of ingroup-outgroup dichotomy. If the attribution of higher F scores to actual lows cannot be explained on the basis of a tendency to avoid commitment by giving mean judgments, there must be some readiness to project differences into the situation whether these differences do or do not exist in fact. In this connection one of the Fiedler, Warrington, and Blaisdell findings (3) is of some relevance. They found that *S*s perceived fellow group members they liked as more similar to themselves than fellow group members they disliked. Since very few lows perceive similarities between their attitudes and those of others, an extrapolation from the Fiedler *et al.* study would contraindicate the presence of an attitude toward people which accepts them openly and trustingly.

Whatever the explanation for the lows' behavior, the present results suggest the need for considerable clarification in the area of the behavior of those who score low on measures of authoritarianism.

SUMMARY

In a previous study it was demonstrated that authoritarians perceived nonauthoritarian peers to have attitudes similar to their own. Nonauthoritarians perceived authoritarians more accurately, but did not estimate authoritarians to be as authoritarian as they actually were. The present study, like the first one, utilized the F scale as a measure of authoritarianism, and extended the earlier study by focussing on the perceptions of authoritarians (high) by other authoritarians and nonauthoritarians (lows) by other nonauthoritarians.

Twelve pairs of highs and 12 pairs of lows were placed in a social situation and instructed to discuss the neutral topics of radio, television, and movies. Each *S* had previously responded to the F scale along with some buffer MMPI items. After the discussion was completed, each *S* was again given the questionnaire, but this time he was instructed to respond to it as he felt the other *S* with whom he had just been speaking would respond to it. The combined data of the two studies led to the following conclusions:

1. Highs tend to estimate peers as having high attitudes, whether these peers are high or low.

2. The estimates by lows of peers are less uniform than those of highs. Most estimates by lows place others in the middle or high range, whether these others are high or low.

The social perceptions of highs appear to be influenced by a "same-stereotype" whereas those of lows are apparently governed by a less extreme and less frequent, but still discernible "other-stereotype." Speculations for this "other-stereotype" are offered.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. CRUTCHFIELD, R. S. Social psychology and group processes. In C. P. Stone & Q. McNemar (Eds.), *Annu. Rev. Psychol.*, 1954, **5**, 171–202.
3. FIEDLER, F. E., WARRINGTON, W. G., & BLAISDELL, F. J. Unconscious attitudes as correlates of sociometric choice in a social group. *J. abnorm. soc. Psychol.*, 1952, **47**, 790–796.
4. SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

Received January 24, 1955.

PHYSIOLOGICAL NEED, SET, AND VISUAL DURATION THRESHOLD

JANET A. TAYLOR
Northwestern University

MANY of the studies purportedly showing the effect of motivation on visual duration thresholds have been criticized on a number of counts, e.g., that word frequency was uncontrolled (8, 14), that threshold differences between groups could be attributed to nonmotivational factors such as individual differences in word usage (14), that the characteristics on which *S*s varied were not necessarily motivational in nature (2), or, in the case of studies of taboo words, that *S*s were reluctant to report their guesses (8, 16). It would appear that such controversies would always be possible whenever acquired characteristics are chosen as the motivational factor to be investigated and used in conjunction with meaningful material. A more clear-cut demonstration of the influence of motivation on visual duration thresholds would seem to be afforded by selecting primary drives such as hunger and thirst, since these are not only agreed upon motivational states but also that differences among *S*s due to past experience which might conceivably affect duration thresholds would presumably be equally distributed between groups. Such a study was performed by Wispé and Drambarean (17) in which it was found that *S*s deprived of food and water for 10 or 24 hours exhibited significantly lower thresholds for need-related words than did control groups tested soon after eating. The differences obtained both for groups shown a list of common or familiar words and for those presented with relatively uncommon words. These results are consistent with those of several other studies measuring the relationship between physiological need and other types of perceptual responses to need-related material (1, 4, 6, 9, 13).

Accepting the finding that higher deprivation levels will be accompanied by lower duration thresholds for need-related words, the present study was undertaken to determine the interaction between physiological need and another variable shown to influence duration thresholds, that of set. A number of studies have shown that when instructions are designed to acquaint *S* with the specific words or type of word to be shown, thresholds for these words tend to decrease (5, 7, 10). The question might be raised as to the effects of introducing instructions informing *S*s, both satiated and deprived, that need-related words will be employed. While the duration thresholds for both groups might be expected to be lower than those exhibited by control *S*s not given such set instructions, deprived *S*s might continue to exhibit lower thresholds than satiated *S*s, or the differences between groups might disappear. The latter result might be taken to indicate that set is a more important variable than physiological need in determining threshold or that need operates to affect performance indirectly through some mediating mechanism like set.

In determining the effect of set on the duration thresholds of groups differing in deprivation level, the present study employed a slightly different procedure than the Wispé and Drambarean study in which each *S* was given a single list containing both neutral and need-related words. Instead *S*s were given two lists of words, the first consisting of neutral words and the second, before which special instructions were introduced for the "set" *S*s, of all need-related words. Such an arrangement permits most of the practice effects to be absorbed by the neutral words and further, might be expected to minimize any self-imposed sets or suspicions about the nature of the experiment that an *S* might have as a consequence of having been asked to control his eating and drinking. By grouping all need-related words into a single list it was hoped that a further aspect of the interaction between need and set could be investigated: whether differences among groups change with successive words on the list as the nature of the words becomes more obvious.

METHOD

Subjects. Two groups of 30 *S*s each were drawn from introductory psychology quiz sections at Northwestern University. The *S*s constituting the deprived groups were volunteers from different sections than those

to which the satiated group belonged. Half of each group was placed in the set condition and half in the no-set condition by alternate assignment as *Ss* appeared to be tested.

Apparatus and materials. Stimulus materials were presented by means of a Gerbrands Mirror Tachistoscope which allows for controlled exposures by means of an automatic timer. The timer was calibrated in steps of .01 sec. and permitted variations between .01 and 1.00 sec.

The stimulus materials consisted of two practice words and two lists of seven words each. The words of the first list (bell, house, story, tail, river, cloud, and post) were neutral in content. The second, need-related, list was composed of the names of foods (meat, cake, bread, steak) and of liquids (milk, water, coffee). All words had a relatively high frequency of usage (50 in one million or above) as determined by the Thorndike-Lorge word count (15).

Procedure. The *Ss* in the deprived groups were tested at 4.30 P.M. and had gone without food or water since breakfast at approximately 7.30 A.M. The *Ss* were thus about 9 hr. deprived. Those in the "satiated" group were asked to eat a normal lunch at the regular time (12 to 12:30 P.M.) and were tested at 1:30 P.M. Each *S* was tested individually.

Preliminary instructions informed *S* that the experiment was part of a larger investigation of the possible influence of degree of deprivation of food and water on different types of perceptual-motor performance. Care was taken not to suggest whether such an influence might be found nor to suggest the direction that it might take.

After the two practice words, the list of seven neutral words was presented, the order of the words being randomized among *Ss*. Following completion of this list, *Ss* were given a brief rest. During this interval additional instructions were given to members of the set subgroups informing them that the postrest words would all refer to the names of food or liquids. The remaining *Ss* were told only that they would be given additional words after the rest. The list of seven need-related words was then presented, the order again being randomized among *Ss*. Guessing was encouraged.

The method of limits, using an ascending series, was employed to determine the threshold for each word. The initial exposure was set at .01 sec. and increased in .01 steps until the word had been correctly identified on two successive exposures. The duration threshold for a word was arbitrarily designated as the duration of the second correct exposure.

RESULTS

Examination of the mean threshold values for the list of neutral words, shown in Table 1, reveals little difference among the means of the groups. The results of an analysis of variance confirm this lack of difference, the between groups F falling far short of statistical significance.

Before attempting to determine whether there was an interaction between the effects of set instructions and deprivation level on the threshold of need-related words, it seemed important to demonstrate that there was a relationship between deprivation level and threshold, as was found in a previous investigation. It was anticipated that *Ss* in the deprived subgroup given no-set instructions would at least initially exhibit lower thresholds than *Ss* in the satiated no-set condition although the difference between the two subgroups might be minimized or disappear later in the list. Inspection of Table 1 shows that this expectation was not confirmed; although t tests indicated no significant differences between the two no-set subgroups, the mean threshold for the deprived *Ss* was actually slightly higher, both for the first word in the list and for the list of need-related words as a whole. While there was no reversal in the means, the performance of the deprived *Ss* in the set condition was also quite close to the corresponding satiated group. The similarity of performance of the deprived and satiated groups is further reflected in the analysis of variance for the total list, summarized in Table 2. The results of this analysis indicate no significant difference between groups and no significant interaction between the set and no-set conditions and deprivation level. Thus, even independent of the set variable, the present results failed to confirm the relationship between physiological need and threshold.

Although no differences were found between the main experimental groups, the set instructions did appear to influence threshold. Thus the mean threshold for each of the two set subgroups was higher than that for each of the no-set groups. The between conditions F, reported in Table 2, was significant at the .05

TABLE 1

MEAN THRESHOLDS (IN SEC.) OF THE TWO GROUPS FOR THE NEUTRAL LIST AND FOR THE NEED-RELATED WORDS IN ORDER OF PRESENTATION

Group	Neutral All	Need Related							
		1	2	3	4	5	6	7	All
Satiated									
No-Set	.071	.075	.067	.071	.069	.062	.069	.061	.068
Set	.076	.065	.065	.065	.070	.067	.062	.065	.066
Total	.074	.070	.066	.068	.070	.065	.065	.063	.067
Deprived									
No-Set	.076	.078	.078	.067	.061	.070	.079	.063	.071
Set	.072	.067	.063	.069	.060	.065	.059	.065	.064
Total	.074	.073	.071	.068	.061	.068	.066	.064	.068

TABLE 2
ANALYSIS OF VARIANCE FOR THRESHOLDS OF NEED-RELATED LIST

Source	*df*	Mean Square	*F*
Groups	1	.26	
Condition	1	1.73	4.12*
Groups × Condition	1	1.01	2.40
Error (within)	56	.42	
Total	59		

* Significant at .05 level.

level. Inspection of the trial by trial data for the two set and the two no-set subgroups indicates that there was some tendency for the difference between the groups to diminish with later words in the list. However, an analysis of variance for trend, comparing the two set subgroups with the two no-set groups, did not yield a significant interaction between conditions and order of words ($F = 2.69$), thus indicating no significant difference in the trend of performance for the set and no-set *S*s.

The experimental design in the present study differed from a previous investigation that indicated a relationship between need and duration threshold, it will be recalled, in that *S*s were given separate lists of neutral and need-related words rather than a single list combining both types of words. Both as a way of determining whether the manner of presentation of the material is a relevant variable and as an additional attempt to demonstrate a relationship between physiological need and threshold, a second experiment was conducted. The second study differed only in that the neutral and need-related words used in the first experiment were presented in random order and that the set variable was omitted.

EXPERIMENT II

Fifteen deprived and 15 satiated *S*s were selected in the same manner as in Experiment I. The procedure for each group was the same as for the no-set groups except for the order of words presented to the *S*. The 14 words from Experiment I were combined into two lists of seven words each, such that three need-related words appeared in one and four in the other. The actual words used in each list and their order of presentation varied in random order from *S* to *S*. The first list given always began with a neutral word. After finishing the first group of seven words, *S* was given a brief rest. Instructions to *S* were the same as the preliminary instructions given in the first study and contained no mention of the presence of the need-related words.

RESULTS AND DISCUSSION

The mean threshold for the deprived group was lower than the threshold for the satiated *S*s on both the neutral and need-related words, the means being .069 vs. .074 sec. in the case of the neutral words and .068 vs. .072 sec. for the need-related. However, the value of Student's *t*, computed to determine the significance of the difference between the means of the need-related words, was only .737. The failure to find differences between groups in the first study, then, cannot reasonably be attributed to the procedure of presenting lists of words of two types.

Unfortunately, the literature concerning motivational variables and perception contains a number of instances in which repetition of an investigation inexplicably fails to duplicate a relationship between the motivational variable and the perceptual response (3, 11, 12). The results of the present study appear to fall into this category. Seemingly, only minor differences—apparatus, actual words employed, etc.—distinguish between the present experiments and those conducted by Wispé and Drambarean. In considering the failure of the present study to reconfirm the relationship between physiological need and duration threshold, it can only be suggested that future studies may find more reliable results or unearth further variables whose uncontrolled effects are responsible for the inconsistencies between the two set of studies.

SUMMARY

Previous research has indicated a relationship between visual duration thresholds and degree of deprivation of food and water. The present study was undertaken to determine whether there was an interaction between the effect of deprivation level and set-inducing instructions on the thresholds of need-related words. Two groups of *S*s, one tested soon after mealtime and the second after approximately 9 hours of food and water deprivation, were presented tachistocscopally with a list of neutral words and then with a list of need-related

words, half of the *Ss* in each group receiving set instructions before the second list. The results indicated that while *Ss* given set instructions exhibited a significantly lower threshold for need-related words than did no-set groups, there was no difference in performance between satiated and deprived *Ss*, even among *Ss* not given set instructions.

A second study was then performed which was similar to the first except that the set variable was omitted and the two types of words were presented in mixed order. Again, no differences were found in the thresholds of the deprived and satiated groups on the need-related words. Thus, the results of two separate experiments failed to duplicate a previously demonstrated relationship between physiological need and duration threshold.

REFERENCES

1. ATKINSON, J. W., & MCCLELLAND, D. C. The projective expression of needs: II. The effect of differential intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–655.
2. BROWN, J. S. Problems presented by the concept of acquired drives. In *Current theory and research in motivation: A symposium*, Lincoln: Univer. of Nebraska Press, 1953.
3. CARTER, L., & SCHOOLER, E. Value, need and other factors in perception, *Psychol. Rev.*, 1949, **56**, 200–208.
4. CHEIN, I., LEVINE, R., & MURPHY, G. The relations of the intensity of a need to the amount of perceptual distortion. *J. Psychol.*, 1942, **13**, 283–293.
5. FREEMAN, J. T. Set or perceptual defense? *J. exp. Psychol.*, 1954, **48**, 283–288.
6. GILCHRIST, J. C., & NESBERG, L. S. Need and perceptual change in need-related objects. *J. exp. Psychol.*, 1952, **44**, 369–376.
7. LACEY, O. W., LEWINGER, H., & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–175.
8. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense," *Psychol. Rev.*, 1950, **57**, 229–234.
9. MCCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205–222.
10. NEISSER, U. An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, 1954, **47**, 399–402.
11. ROCK, I., & FLECK, F. S. A re-examination of the effect of monetary reward and punishment on figure-ground perception. *J. exp. Psychol.*, 1950, **40**, 766–776.
12. ROSENTHAL, B. G. Attitude towards money, need, and methods of presentation as determinants of perception of coins from six to ten years of age. *Amer. Psychologist*, 1951, **6**, 317. (Abstract)
13. SANFORD, R. N. The effects of the abstinence from food upon the imaginal processes. A further study. *J. soc. Psychol.*, 1937, **8**, 335–343.
14. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.
15. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.
16. WHITTAKER, EDNA M., GILCHRIST, J. C., & FISCHER, JEAN W. Perceptual defense or response suppression. *J. abnorm. soc. Psychol.*, 1952, **47**, 732–733.
17. WISPÉ, L. G., & DRAMBAREAN, N. C. Physiological need, word frequency, and visual duration response. *J. exp. Psychol.*, 1953, **46**, 25–31.

Received January 24, 1955.

THE REINFORCEMENT OF COOPERATION BETWEEN CHILDREN[1]

NATHAN H. AZRIN
Harvard University

AND

OGDEN R. LINDSLEY
Harvard Medical School

Most methods for the development and experimental analysis of cooperation between humans require specific instructions concerning the cooperative relationship between the individual responses. Peters and Murphree have developed one of the most recent of these methods (1). Skinner has suggested (2), and shown with lower organisms (3), that cooperation between individuals can be developed, maintained, and eliminated solely by manipulating the contingency between reinforcing stimuli and the cooperative response.

The advantages of eliminating instructions concerning cooperation are that (*a*) the initial acquisition of cooperation can be studied, (*b*) subjects (*S*s) that learn by demonstration and instruction with difficulty (i.e., infants, certain classes of psychotics, and lower organisms) can be studied, and (*c*) no problems involving the effects of instructions upon the behavior of the *S*s are involved.

Some more general advantages of operant conditioning techniques are (*a*) a more continuous record of the cooperative process is obtained, (*b*) extraneous environmental variables are minimized, and (*c*) relatively long periods of experimental observation are possible.

Problem

Can cooperation between children be developed, maintained, and eliminated solely by the presentation or nonpresentation of a single reinforcing stimulus, available to each member of the cooperative team, following each cooperative response?

Cooperative Teams

Twenty children, seven to twelve years of age, were formed into ten cooperative teams of two children. The children in each team were matched as to age and sex. Seven teams were boys and three were girls.[2] Selection was made via the request, "Who wants to play a game?" The first two volunteers of the same age and sex were chosen for each team. The age given by the children was verified against available community center records. No information concerning the game was given during the selection. No teams were rejected.

Cooperative Response

Cooperation was assured by designing an apparatus that (*a*) could not be operated by one individual alone (assuring group behavior), and (*b*) demanded that one individual respond to the behavior of the other individual in order to produce reinforcement (assuring cooperation).

Procedure

The two children of each cooperative team were placed at opposite sides of a table with three holes and a stylus in front of each child (see Fig. 1). A wire screen down the center of the table prevented each child from manipulating the other child's stylus, which was on the other side of the table.

The following instructions were given: "This is a game. You can play the game any way you want to or do anything else that you want to do. This is how the game works: Put both sticks (styli) into all three of the holes." (This sentence was repeated until both styli had been placed in the three available holes.) "While you are in this room some of these" (the experimenter (*E*) held out several jelly beans) "will drop into this cup. You can eat them here if you want to or you can take them home with you." The instructions were then repeated without reply to any questions, after which *E* said: "I am leaving the room now; you can play any game that you want to while I am gone." Then *E* left the room until the end of the experimental session.

If the styli were placed in opposite holes within 0.04 seconds of each other (a cooperative response), a red light flashed on the table (conditioned reinforcing stimulus) and a single jelly bean (reinforcing stimulus) fell into the cup that was accessible to both children.[3] Cooperative responses were recorded on counters and a cumulative response recorder in an adjoining room.

Experimental Design

Each team was studied for one continuous experimental session divided into the following three consecutive periods without experimental interruption:

1. ***First reinforcement period.*** Every cooperative response was reinforced for over 15 min. If the rate of response was not steady at this time, the reinforcement was continued until five minutes passed with no noticeable change in the rate of cooperation.

2. ***Extinction period.*** The cooperative responses were not reinforced for a period of at least 15 minutes and until a steady rate of response for at least five minutes was observed.

3. ***Second reinforcement period.*** The cooperative responses were again reinforced until at least three min-

[1] This paper was read at a meeting of the Eastern Psychological Association on April 10, 1954, New York City.

[2] We wish to thank the Harriet Tubman House and the South Bay Union of Boston, Mass., for providing the subjects and the use of their facilities.

[3] Skinner (3) presented two reinforcing stimuli (one to each pigeon) following each cooperative response.

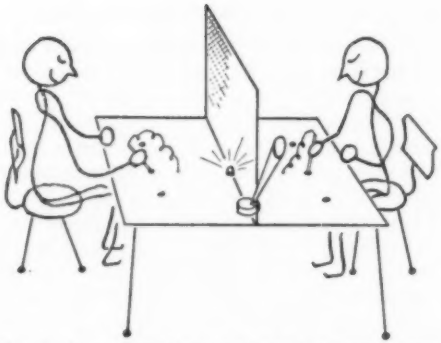


FIG. 1. APPARATUS USED FOR THE REINFORCEMENT OF COOPERATION BETWEEN CHILDREN

utes of a stable rate occurred. This was done to determine whether a reduction in rate during the extinction period was due to extinction, satiation, or fatigue.

RESULTS

All teams learned to cooperate without specific instructions in the first 10 min. of experimentation. Observation through a one-way vision screen disclosed that leader-follower relationships were developed and maintained in most cases. Almost immediately eight teams divided the candy in some manner. With two teams, one member at first took all the candy until the other member refused to cooperate. When verbal agreement was reached in these two teams, the members then cooperated and divided the candy. Most vocalization occurred during the initial acquisition period and throughout the extinction period. This vocalization was correlated with a higher variability in rate during these periods. (See below.)

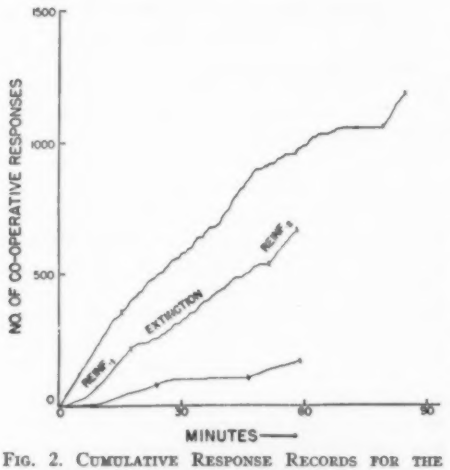


FIG. 2. CUMULATIVE RESPONSE RECORDS FOR THE TEAMS WITH THE HIGHEST, MEDIAN, AND LOWEST RATES OF COOPERATION

TABLE 1

THE MEDIAN AND RANGE OF THE NUMBER OF COOPERATIVE RESPONSES PER MINUTE FOR THE CRITICAL EXPERIMENTAL PERIODS

N 10	Number of cooperative responses per minute			
	First three mins. of first reinf. period.	Last three mins. of first reinf. period.	Last three mins. of extinction period.	Last three mins. of second reinf. period.
Median	5.5	17.5	1.5	17.5
Range	1–26	6–26	0–7	6–27

Figure 2 contains cumulative records of the cooperative responses of the three teams with the highest, the median, and the lowest number of cooperative responses for the experimental session. These curves show a large difference in the rate of acquisition of cooperation. One team took almost 10 minutes to acquire a high cooperative response rate. Stable rates of cooperation can be observed during the latter parts of the first reinforcement period. The gradual, rather than immediate, decline in cooperation during extinction suggests an orderly extinction of cooperative behavior as is found with individual extinction curves. In all cases the variability of rate was greater during extinction than during reinforcement. Skinner has found this increased variability in rate during extinction with lower organisms and has described it as emotional behavior (2, p. 69). The high rate of response following the first reinforcement of the second reinforcement period shows that reacquisition is almost immediate.

Table 1 contains a quantification of the records for statistical analysis. The median and range of the number of cooperative responses per minute for all 10 teams during the critical periods of the experiment are given. The number of cooperative responses per minute for the first three minutes of the first reinforcement period was significantly lower than the rate during the last three minutes of the first reinforcement period ($p < .02$).[4] This shows that

[4] Wilcoxon's nonparametric T for paired associates was used in all statistical treatments (4).

the rate of cooperation was significantly lower during initial acquisition than during maintenance of cooperation. The number of cooperative responses per minute during the last three minutes of extinction was significantly lower than the rate during the last three minutes of the first reinforcement period ($p < .001$). This shows that the removal of reinforcement during extinction significantly lowered the rate of cooperation between these children.

The number of cooperative responses per minute during the last three minutes of the second reinforcement period was significantly above the rate during the last three minutes of the extinction period ($p < .001$). This shows that the rate of cooperation was significantly increased during the second reinforcement period and that the drop in rate during extinction was due to the absence of the reinforcing stimulus rather than satiation or fatigue. The rates of cooperation during the second reinforcement period and the last three minutes of the first reinforcement period were not significantly different and show that the rate was almost immediately restored to its pre-extinction value upon the presentation of reinforcement for the second time. The rate of cooperative responding during the first three minutes of the second reinforcement period was significantly higher than during the first three minutes of the first reinforcement period ($p < .02$). This again shows that the reacquisition of cooperation was not gradual, as was initial acquisition, but occurred almost immediately.

CONCLUSIONS

Operant conditioning techniques can be used to develop, maintain, and eliminate cooperation between children without the use of specific instructions concerning cooperation. The rate of a cooperative response changes in much the same way as a function of single reinforcements as does an individual response. In the reinforcement of cooperative responses, a reinforcing stimulus need not be delivered to each member of the cooperative team following each cooperative response. The presentation of a single reinforcing stimulus, available to each member of the cooperative team, is sufficient to increase the rate of cooperation. The cooperative response gradually increases in frequency when reinforced and gradually decreases in frequency when no longer reinforced (extinction). Cooperative responses are maintained at a stable rate during reinforcement but occur in sporadic bursts during extinction. Reinforcement following extinction results in an almost immediate restoration of the rate of cooperation to its pre-extinction value.

REFERENCES

1. PETERS, H. N., & MURPHREE, O. D. A cooperative multiple-choice apparatus. *Science*, 1954, **119**, 189–191.
2. SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
3. SKINNER, B. F. Classroom demonstration. Personal communication, 1952.
4. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.

Received January 27, 1955.

THE EFFECTS OF AN "ANXIETY-REDUCING" MEDICATION ON GROUP BEHAVIOR UNDER THREAT[1]

JOHN T. LANZETTA
Air Force Personnel and Training Research Center, Randolph Air Force Base

AND

G. R. WENDT, PETER LANGHAM, AND DON HAEFNER
University of Rochester

A PREVIOUS report has presented the results of a study on group behavior under conditions of situational threat (10). The findings indicated that threat influenced behavior in primarily three areas of functioning: there were differences in the nature of interpersonal affective behavior, in the characteristic ways in which participants utilized their own and other group members' resources, and in the effectiveness of the groups' task performance. Threat resulted in a reduction of forceful, assertive, interpersonal, as well as task-directed behavior, and to increased concern with group acceptance. Threat groups were also less effective and showed greater variability in effectiveness.

To account for these and similar results reported by Wright (16) and Lanzetta (9), it was postulated that the primary concomitant of threat was an increase in anxiety and, further, that an increase in anxiety accentuated "security" responses in an attempt to terminate or diminish the anxiety. Since the potential support of other individuals was available, behavior most likely to assure this support was elicited. Individuals inhibited behavior that would alienate them from, or cause isolation by, the other group members.

The primary role assigned to anxiety in this formulation emphasized the necessity for independent manipulation of this variable. If an increase in anxiety due to the situational threat was responsible for the behavioral differences noted, then change in "anxiety level" would influence the effects of threat on behavior. One possible means of affecting the anxiety level of individuals without fundamentally changing the situational factors responsible for the imposition of threat is by the use of chemical agents. Masserman (12), and Miller (4) have used sodium amytal in studies of fear; and Wendt, *et al.* (7), have explored the effects of several anxiety-modifying chemical agents on social behavior. The general role of barbiturates as anxiety reducers can also be inferred from clinical case histories (6). Since present evidence, though, does not allow unequivocal prediction that a particular agent will have a specific, circumscribed effect on motivational and/or emotional states, their use to test hypotheses is hardly tenable. The effect of the agent can be inferred only from responses, and in the early stages of investigation there is some tendency toward circularity.

In the present study, interest was in the effects of a chemical agent on behavior under conditions of situational threat. The agent chosen was one that resulted in changes in behavioral and self-report responses, which one might infer were a result of a decrease in "anxiety level" (7). Since the evidence upon which such an inference could be based was not extensive, the study was considered an exploratory one.

METHOD

Subjects and design. The main independent variables were the threat (Th.) and threat-medicated (Th. Med.) treatments.[2] Subjects (*Ss*) for the experimental treatments were 168 Air Force and Navy Reserve Officer Trainee Candidates (ROTC) in freshman and sophomore classes at the University of Rochester. They comprised about 95 per cent of the total enrollment of these groups. The 168 *Ss* were randomly assigned to groups of three and then randomly assigned to treatments. Thirty-six groups were assigned to the Th. and 20 groups to the Th.-Med. conditions.

The *Ss* for a control condition (C) were 27 volunteers from the freshman and sophomore non-ROTC student

[1] The work described in this paper was supported by funds from contracts between the University of Rochester and the Office of Naval Research: Contracts N-60nr-241 and N60ri-126. The latter contract was under the general direction of Dr. G. R. Wendt.

[2] Within the Th. and Th.-Med. treatments, a distinction was made between "locus" and "target" of threat, each being dichotomized: locus into external or internal to the group, and target into group- or individual-directed. A 2 x 2 factorial design was used within each treatment with nine groups randomly assigned to each condition under Th. and five groups to each condition under the Th.-Med. treatment. No significant variance on any of the measures could be attributed to conditions within treatment; thus all groups within a treatment were combined for the present analysis.

population. They were randomly assigned to nine groups of three members each. A Th. placebo-control condition was not felt to be necessary since previous work with the medication used revealed nonsignificant placebo-control differences with respect to social behavior variables similar to those under study in the present experiment (7).

Experimental treatments. Threat was produced by structuring the group sessions as evaluation tests. Previous laboratory studies of leadership had utilized ROTC *S*s, and to obtain criterion data for one of these studies *S*s were evaluated while on their summer cruise. Since that time there had been inquiries by other ROTC students as to the results of this evaluation and as to whether a similar procedure might again be utilized. This orientation toward the project staff, as an evaluative group, was utilized to create the threat.

Experimental *S*s were informed by *E*, with the active support of their ROTC commanding officers, that they were to be evaluated. They were told that the results of this assessment were to become part of their permanent record and were to be used by the local units for general evaluation purposes. The specific function of the reports was deliberately left ambiguous on the assumption that this would be more "threatening." This "evaluation threat" was reinforced by several impersonal letters and rather curt scheduling notices, as well as a formal attitude on the part of the experimenter. The assumption that such a procedure would be threatening seems warranted by the behavior of *S*s when they reported to the experimental room. In general they were tense, subdued, and rather passive; *E* had to assume the responsibility for getting them to "look around." By contrast, the control *S*s often had to be restrained from too extensive exploration of the experimental room and manipulation of task materials.

Th.-Med. groups were randomly selected from the sample that had undergone this recruiting and indoctrination procedure. In addition, they were required to ingest orally .5 mg. of a mixture of Seconal and Benzedrine in capsule form. The capsule was administered two hours before the *S*s were to report for assessment. The *S*s reported to a different building from that in which the experimental rooms were located, received the medication under the observation of an assistant, and then resumed their regular classroom schedule until the time to report for the evaluation tests.

Introduction of the medication for the Th.-Med. treatment without raising suspicion as to the purposes of the situational tests proved a difficult problem. A request from a project engaged in medication research was made to these groups for assistance in the testing of a motion-sickness preventive. The *S*s were informed the preventive had been pretested with a small sample of subjects and no effects had been reported; additional tests were necessary only to satisfy military standards for accepting the medication for operational use.[3] The complete indoctrination also presented the background and importance of research on motion-sickness preventives and emphasized the minimum effort necessary since *S*s were already required to report for evaluation.

About 5 per cent of the *S*s refused the request. In such cases a volunteer who had been similarly recruited and indoctrinated was substituted in the groups and the "refusal" was placed in a group in the Th. treatment. Violation of strict random assignment was thus necessary, although the small percentage of refusals kept it to a minimum.

Subjects under control conditions were instructed by a different *E*. They were informed they were being used for a study necessary to complete requirements for a graduate course. The *E* attempted to set an informal, relaxed atmosphere, and usually spent five to ten minutes interacting with them before the session began.

Task[4]. The task utilized, called "Intercept," was a strategy game which required rapid adaptation to changing conditions and continued interdependence between members in reaching group decisions. Each member had a certain portion of the information necessary for group decision, and each was exposed to conflicting pressures to act in his own as well as the groups interest. A description of the task and its administration can be found in a previous report (10).

Instruments. Measures of overt behavior were obtained by the use of a technique for recording on-going interaction and by ratings made at the end of the experimental session. Four observers watched each group session through a one-way-vision mirror. Two observers recorded the social interaction of each *S* in terms of an adaptation of the system proposed by Bales (1). The system was elaborated by the use of subcategories used by Bales in defining his main categories. A full description of the system is presented by Bales (1, pp. 177–195). In addition, four observers rated each *S* at the end of the session in respect to his cooperation, efficiency, leadership, aggression, autocracy, sociability, adaptability, achievement motivation, and the extent to which he was self-oriented or oriented toward group acceptance. The ratings were derived from previous factor analytic studies of rating variables (2, 3) and from studies on decision-making groups (11).

An adjective check list that had proved suitable for measuring changes in attitudes and motivation that follow medication was used for obtaining self-report data. Subjects read quickly through a list of adjectives, checking those that they felt were descriptive of them at the moment. The list consisted of 114 adjectives, defining 12 dimensions. A full description of its development is presented by Wendt, *et al.* (7) and Nowlis (14).

RESULTS

Available for analysis were ratings (by four observers) of each individual on 10 traits, a record of the on-going behavior of each group member in terms of Bales's 12 behavior categories, and subjects' responses to an adjective check list. The average reliabilities for the observer ratings and categorizations were .77 and .60, respectively.

[3] G. R. Wendt and his associates have been conducting motion-sickness research for a considerable period of years, and the elevator and rotating chair are familiar to most students on the campus. Thus a request for assistance from this group would not appear unusual.

[4] The task is a modification of one developed by W. C. Schutz. A more detailed discussion can be found in his report (15).

In order to determine whether differences due to experimental treatments were significant, *t*'s between the threat (Th.), threat-medicated (Th.-Med.), and control (C) conditions were computed. In those cases where the *F* ratio indicated a significant difference in variance between the two samples, Cochran and Cox's correction, as given in Edwards (5, p. 168), was used in evaluating the significance of the obtained *t*. The results of these tests, by instrument, are presented in Tables 1, 2, and 3.

As an examination of Table 1 indicates, significant differences were obtained between Th. and Th.-Med. conditions on four of the indices of the adjective check list. Th.-Med. *S*s report more organic symptoms, and report being more elated, relaxed, and cooperative than Th. groups. Th.-Med. groups differ significantly from control groups only in the greater number of organic symptoms reported.

Although other differences are nonsignificant, there is considerable consistency in the reports of the treatment groups. Th.-Med.

TABLE 1

MEAN DIFFERENCES IN ADJECTIVE CHECK-LIST INDEX SCORES FOR THE THREE EXPERIMENTAL TREATMENTS

	Index	Th.—Th. Med. Md	Th.—Th. Med. σ md	C—Th. Med. Md	C—Th. Med. σ md	Th.—C Md	Th.—C σ md
I	Anxiety	.79	.91	.76	1.33	.03	1.38
IIa	Elation	−1.95**	.94	−.62	−1.80	−1.33	1.78
IIb	Depression	.77	.49	.02	.53	.75	.59
III	Hostile Aggression	.22	.68	1.05	1.02	−.83	1.07
IV	Organic Symptoms	−2.59**	1.03	−3.70***	1.02	1.11**	.47
V	Relaxation	−2.35**	1.08	−1.52	1.77	−.83	1.60
VI	Positive Aggression	−1.11	2.09	.75	2.69	−1.86	2.26
VIIa	Cooperation	−3.58*	1.86	−2.86	2.59	−.72	2.21
VIIb	Obstruction	.60	.80	.41	1.22	.19	1.21
VIIIa	Dominance	−1.80	1.69	−1.05	2.08	−.75	1.74
VIIIb	Submission	−.93	.86	−1.40	.82	.47	.77
IX	Task Involvement	−1.78	2.56	−1.26	3.26	−.52	2.74

Note.—N = 36 for Th., 20 for Th.-Med., and 9 for C treatments.
* .10 > p > .05.
** .05 > p > .01.
*** .01 > p.

TABLE 2

MEAN DIFFERENCES IN THE PROPORTION OF ACTS IN EACH INTERACTION CATEGORY FOR THE THREE EXPERIMENTAL TREATMENTS

	Category	Th.—Th. Med. Md	Th.—Th. Med. σ md	C—Th. Med. Md	C—Th. Med. σ md	Th.—C Md	Th.—C σ md
I	Shows Solidarity	.06	.19	.32	.27	−.25	.28
II	Tension Release	.35***	.12	.35	.25	.00	—
III	Agrees	−.44	.43	.02	.64	−.46	.63
IV	Gives Suggestions	−1.67***	.55	.49	.69	−2.16***	.64
V	Gives Opinions	−.91	1.48	.08	1.84	−.99	1.55
VI	Gives Orientation	.17	1.15	−1.25	1.42	1.42	1.25
VII	Asks Orientation	.48	.41	−.66	.48	.71***	.14
VIII	Asks Opinions	−.57	.41	−1.39**	.65	.82	.54
IX	Asks Suggestions	.13	.11	.18	.17	−.05	.17
X	Disagrees	1.05***	.39	.15	.53	.90	.62
XI	Shows Tension	1.08***	.37	1.07	.74	−.38	.76
XII	Shows Antagonism	.67*	.37	1.07	.74	−.38	.76
	Total Activity	−118.42*	45.16	17.74	48.90	−154.75***	41.01

Note.—N = 36 for Th., 20 for Th.-Med., and 9 for C treatments.
* .10 > p > .05.
** .05 > p > .01.
*** .01 > p.

TABLE 3

MEAN DIFFERENCES IN RATINGS (SUMMED OVER OBSERVERS) FOR THE THREE EXPERIMENTAL TREATMENTS

Rating	Th.—Th. Med. Md	Th.—Th. Med. σ md	C—Th. Med. Md	C—Th. Med. σ md	Th.—C Md	Th.—C σ md
1. Cooperation	−.32	1.98	−3.08	2.72	2.56	2.64
2. Efficiency	1.75	2.79	6.00**	2.77	−4.25**	1.84
3. Leadership	1.16	1.97	3.71*	1.92	−2.55	1.45
4. Aggression	1.61	3.26	8.80	4.94	−7.19	4.50
5. Autocracy	.56	2.58	7.70*	3.92	−7.14	3.80
6. Sociability	3.12	2.14	1.76	3.70	1.36	3.80
7. Striving for Group Acceptance	.97	1.59	−2.75	1.57	3.72**	1.54
8. Adaptability	.78	2.36	4.39**	2.02	−3.81**	1.42
9. Self-Orientedness	3.68	2.45	3.32	3.34	.36	3.20
10. Achievement Motivation	1.33	2.73	6.56*	3.41	−5.23	2.81

Note.—$N = 36$ for Th., 20 for Th.-Med., and 9 for C treatments.
* $.10 > p > .05$.
** $.05 > p > .01$.
*** $.01 > p$.

groups check the most adjectives connoting a relaxed, positively affective, cooperative, phenomenal state, and the least adjectives indicating an anxious, aggressive, obstructive state. Th. groups report more anxiety and depression than either of the other groups. If self reports can be considered valid criteria, the experimental treatments did produce changes in the emotional and/or motivational state of *S*s. Change is evident along three dimensions: anxiety, elation, and aggression or assertion. The conditions can be ordered along these dimensions as follows:

a. anxiety—Th. > C > Th.-Med.
b. elation—Th.-Med. > C > Th.
c. assertion—C > Th. > Th.-Med.

Behavioral differences for the treatment groups are consistent with these results. Table 2 presents the mean differences between the conditions for the interaction categories, while Table 3 shows similar data for the observer ratings. Significant differences were obtained between Th. and Th.-Med. treatments for five of the category indices (Table 2). Th.-Med. groups showed less tension and tension release, disagreed to a lesser extent, and were less antagonistic. They also gave more suggestions and were more active.

The results for the observer ratings (Table 3) indicate no significant differences for the Th.—Th.-Med. treatments. Th.-Med. groups, though, are rated as less efficient, less autocratic, less adaptable, and less achievement motivated than Control groups. They also received significantly lower ratings on leadership.

The category system, apparently, was sensitive to differences between the Th. and Th.-Med. treatments, while the ratings were sensitive to differences between the Th.-Med. and C conditions. The differential discrimination of the two techniques appears inexplicable. Though ex post facto explanations are always hazardous, one appears especially likely. Ratings are of a more global nature and, in general, require inferences about the motives or consequents of behavior. For example, ratings of leadership or efficiency are based on integration and collation of a large number of behavioral events and require judgments about the effects or consequences of the behavior. The category system, on the other hand, requires judgments about the placement of *acts* within specific categories. The criteria for inclusion do not, in general, require inferences about motivation or consequents. Thus, two individuals could be equally active, ask a similar number of questions, and give an equal number of suggestions, and still be rated quite differently on a number of traits. If one's suggestions were more pertinent, if his questions more frequently concerned task problems, and his activity was more often task oriented, he would be rated higher on achievement motivation, efficiency, and possibly leadership.

In a sense, then, the categories reflect the interactive, communicative processes in the group, the ratings the motives underlying, and the consequences of, this interaction. Thus one would infer that medicated *S*s under threat interact in essentially similar fashions as C

groups. That is, they ask as many questions, give as much information, etc. Apparently, though, this behavior is not especially task oriented, as evidenced by the ratings. Medicated groups might be characterized as having an active, nonaggressive "good time" with no especial concern for effective performance on the task.

DISCUSSION

In view of the significant behavioral changes previously reported for the threat treatment (10), it is apparent that the medication employed did affect responses to threat. Threat-Medicated subjects differed from controls in several important respects, but, in general, responded more like the control *S*s than like the threat *S*s. Apparently the medication served to reduce considerably the negative emotional concomitants of the threat treatment, resulting also in mild organic symptoms.

In attempting to understand the results, we can use as a starting point the inferred change in the emotional and motivational states due to the experimental treatments. Medication resulted in a marked decrease in anxiety such that on some arbitrary anxiety scale our conditions might be ordered as follows: Threat > Control > Threat-Medicated. The reduction in anxiety and associated elation resulted in increased activity, but of a nontask-oriented, nonaggressive nature. If one assumes that a certain minimum level of anxiety is necessary for effective task orientation and performance, then the medication seemed to cause a reduction in anxiety below this point. In view of the decreased effectiveness of the threat groups also, what must be postulated is a curvilinear relationship between anxiety level and effectiveness.

The relationship is analogous to that assumed to hold between anxiety level and "socialized" behavior. It is typically postulated that some minimum level of anxiety is necessary for normal social development; marked increases in anxiety tend to lead to behavior pathology, while marked reduction in anxiety (lobotomized patients, psychopathic behavior disorders) results in arrested social development (8, 13).

Performance under C conditions is thus assumed to involve some measure of anxiety associated with the possibility of task failure. Threat increases this anxiety to a level sufficient to elicit interfering "security-oriented" responses, while medication reduces the anxiety to a level where normal inhibitions and concern with achievement are decreased.

SUMMARY

Three-member teams worked on a group task under three experimental conditions: a threat condition in which *S*s received an "anxiety-reducing" medication (Th.-Med.), a threat condition with *S*s receiving no medication (Th.), and a control (nonthreat) condition (C). Threat was imposed by structuring the sessions as evaluation tests whose results would become part of the *S*s' ROTC records.

Behavioral indices were obtained from observers' categorizations of the behavior of group members using Bales's category system, and ratings of each member on 10 characteristics. The *S*s' perceptions of their group and their own motivational state were obtained from responses to an adjective check list.

The results indicate that the experimental treatments did produce changes in the emotional and/or motivational state of *S*s. Change was evident along three dimensions as follows:

a. anxiety—Th. > C > Th.-Med.
b. elation—Th.-Med. > C > Th.
c. assertion—C > Th. > Th.-Med.

Behavioral differences for the treatment groups were consistent with these results. Th.-Med. groups showed less tension and tension release, disagreed to a lesser extent, were less antagonistic, and were more active than Th. groups. Th.-Med. groups, though, were rated as less efficient, less autocratic, less adaptable, and less achievement motivated than Control groups. They also received significantly lower ratings on leadership.

Apparently medicated *S*s under threat interact in essentially similar fashions as Control groups that is, they ask as many questions, give as much information, etc. Their behavior, though, is not task oriented. Medicated groups may be characterized as having an active, nonaggressive "good time" with no especial concern for effective performance on the task.

REFERENCES

1. BALES, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley, 1950.
2. CARTER, L. F. Some research on leadership in small

groups. In H. Guetzkow (Ed.) *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951, pp. 146–157.

3. COUCH, A., & CARTER, L. F. A factorial study of the rated behavior of group members. Paper presented at East. Psychol. Ass., Brooklyn, April 1952.
4. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
5. EDWARDS, A. L. *Experimental design in psychological research,* New York: Rinehart, 1950.
6. GRINKER, R. R., & SPEIGEL, J. P. *War neurosis.* New York: Blakiston, 1945.
7. HARWAY, V. T., LANZETTA, J. T., NOWLIS, HELEN H., NOWLIS, V., RIESEN, A. H., & WENDT, G. R. Chemical influences on behavior, II. Rochester, N. Y.: Univer. of Rochester, 1953. (Tech. Rep., Project No. NR144-060.)
8. HOCH, P. H., & ZUBIN, J. (Eds.) *Anxiety.* New York: Grune & Stratton, 1950.
9. LANZETTA, J. T. Group behavior under stress. *Hum. Relat.,* in press.
10. LANZETTA, J. T., HAEFNER, D., LANGHAM, P., & AXELROD, H. Some effects of situational threat on group behavior. *J. abnorm. soc. Psychol.,* 1954, **3,** 445–453.
11. MARQUIS, D. G., GUETZKOW, H., & HEYNS, R. W. A social psychological study of the decision making conference. In H. Guetzkow (Ed.) *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951, pp. 55–67.
12. MASSERMAN, J. H. *Behavior and neurosis.* Chicago: Univer. of Chicago Press, 1953.
13. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
14. NOWLIS, V. The development and modification of motivational systems in personality. In *Current theory and research in motivation.* Lincoln: Univer. of Nebraska Press, 1953, pp. 114–138.
15. SCHUTZ, W. C. Studies in group behavior: I. Construction of high productivity groups. Depart. of Systems Analysis, Tufts Coll., 1953. (Rep. No. 1953-494-03-04.)
16. WRIGHT, M. E. The influence of frustration upon the social relations of young children. *Charact. & Pers.,* 1943, **12,** 111–122.

Receive February 1, 1955.

THE EFFECT OF INDUCED AGGRESSIVENESS ON OPINION CHANGE[1]

WALTER WEISS AND BERNARD J. FINE

Boston University

PREVIOUS research by Weiss and Fine (14) has indicated that subjects evidencing high aggressiveness, as measured by the TAT and the Rosenzweig P-F Study, were influenced more by a communication designed to provoke punitive opinions than were those evidencing low fantasy aggressiveness. This finding supported the data of earlier correlational research which related personality variables to opinions at one time, but not to *changes* in opinions (1, 2). The general research hypothesis for the authors' previous study presented a compound independent variable involving aggressiveness and exposure to a communication. Only the latter aspect was experimentally controlled. The theoretical advantage of verifying an hypothesis by complete experimental methodology was the primary motive for the present research. While personality factors, such as aggressiveness, can not readily be influenced, temporary states analogous to them can be induced. This was the method adopted for the current investigation.

The broad theoretical statement underlying the research asserts that communication effectiveness is facilitated when congruency exists between response predispositions of the communicatees and the communication appeals. To elaborate this interactive relationship, the effect of a communication designed to induce lenient opinions on a topic should be examined. The prediction would then be that high aggressives would be less influenced by such a communication than would low aggressives. The verification of this prediction was the second purpose of the current research. It was also considered advisable in view of the assumption of congruency to attempt the development of an ego-satisfying experience, so that aggression arousal may be contrasted with ego satisfaction rather than simply no aggression arousal. As a minimum, this second type of induction should aid the production of two clearly separable groups along the dimension of aggressiveness.

[1] The research reported in this article was done under Contract NONR-492(04), between Boston University and the Office of Naval Research.

METHOD

General design. A before-after design was employed. The before-measure of opinions on juvenile delinquency and on the relations of the United States with her allies was taken by the subjects (*Ss*) in one of their classes. The Rosenzweig P-F Study was also completed at the same time. Approximately two weeks later, *Ss* were involved individually in either a frustrating, aggression-arousing situation or in an ego-satisfying situation. As part of the experimental session, they read an article on delinquency or on the relations of our allies with communist China. Immediately afterwards, they responded to a questionnaire seeking their reactions to the article they had read and their opinions on a variety of topics.

Subjects. Ninety-six students from courses in journalism, radio programming, and introductory psychology at Boston University, who volunteered for the individual sessions, comprise the sample employed. The *Ss* were paid for the experimental hour.

Before-measures. During a meeting of their respective courses, the students filled out an opinion questionnaire containing four items on juvenile delinquency and four on America's relations with her allies. All eight items sought responses on policy; e.g., should delinquents be treated leniently or punitively, or should America continue aid to allies who do not follow our lead vis-a-vis communist China. The questionnaire was presented as part of a survey of opinions of college students being conducted by the Division of Research of Boston University. Following this, the students completed the adult form of the Rosenzweig P-F Study, which was ostensibly being tested by the Division as a measure of "creative imagination." Then paid volunteers were requested for an individual session to determine the value of several other tests of creative imagination.

Experimental session. The two communications and manipulated atmospheres were ordered factorially to yield a fourfold arrangement of treatments. Thus, 48 *Ss* were exposed to the frustrating, aggression-arousing conditions, and 48 to the ego-satisfying treatment; then half of each group read the communication on delinquency and half the communication on the relations of our allies with communist China. A replication of the basic two-by-two design was completed every time a set of four *Ss* was tested. Within each replication, *Ss* were assigned at random to one of the four treatment combinations. A different random assignment was employed for each replication.

At the beginning of an experimental session, *Ss* were informed that, although designed primarily to measure creative imagination, the tests do provide indirectly an indication of intellectual capacity. The students were also informed that half of the *Ss* would be asked to return at a later time to take other tests and that the

criterion for selection of *Ss* would be general cooperativeness with the *E* during the first hour. The instructions were then given for the Arthur Stencil Design Test I. As part of the instructions, the *Ss* were told that time norms for completion of the designs were available for college students; thus they would be informed by *E* at the completion of each problem of the adequacy of their performance in comparison with the norm. Furthermore, a maximum of one minute would be allowed for completing each of the designs; this they were told was more than sufficient time for the vast majority of college students. Only eight design problems from the available set of 20 were presented to the *Ss*. The first three were the same for all *Ss*. They were relatively easy and were completed within a minute by all. The remaining five designs for the *Ss* involved in the ego-satisfying situation were somewhat more complex than the first three but still sufficiently simple so that over 90 per cent of these *Ss* were able to complete them. They were also advised that their times for completion were better than the average. Furthermore, *E* made personally complimentary and encouraging remarks to them. The frustrated, aggression-aroused *Ss* were given five designs of greater complexity, whose average times for completion was determined by pretest to be beyond one minute. Most of these *Ss* failed at least three out of five of the complex designs; they were told that the time scores, for those they were able to complete, were beyond the norm. Also *E* made uncomplimentary and somewhat insulting comments about their performance and abilities, which increased in intensity at the end of the Stencil Design Test. The content and intensity of the insulting remarks varied slightly from *S* to *S* depending on their observed effect and on the known academic aspirations of the student.

Following the Design Test, *Ss* were asked to write stories for two TAT cards, 18GF and 3BM. These were selected as potential discriminators between aggressive and nonaggressive *Ss* on the basis of previous research (7, 14). The abbreviated TAT was presented as a second test of creative imagination. The *Ss* were allowed five minutes to write each story on a page of a prepared booklet. The *E* examined and commented upon each story immediately after it was completed. Complimentary remarks were made about the stories of the nonfrustrated *Ss*. The intensity of the derogatory, insulting comments was increased for the other *Ss*.

At the completion of the TAT, *E* told the *Ss* that for the final test they would be asked to read carefully an article and then to respond to some questions about their reactions to it. He advised the nonfrustrated *Ss* to continue to do well and the frustrated *Ss* to improve their performance, for it was the over-all impression on all three tests of creative imagination that was important. All *Ss* were then informed that a second *E* would administer this last measure.[2]

The article was presented as one received for publication by the editors of a national magazine. We, as well as the editors, were interested in determining if college students and professional editors could arrive at the same judgments of an article by an unnamed author. The *Ss* were informed that 10 minutes would be allowed for reading the article, after which they would be asked questions about their reactions to it. Depending on which treatment combination the *S* had been assigned to, he read either the article on delinquency or the one on the relations of our allies with communist China. The former argued strongly for adopting punitive measures to curb juvenile offenses in contrast to practices of leniency and guidance. The latter urged Americans to consider the special economic, geographic, and military circumstances of other nations which incline them to adopt policies toward China on such matters as trade and diplomatic recognition which differ from our own; also the impracticality of diplomatic threat and withdrawal of aid as a method to compel free nations to follow our lead was emphasized.

Immediately after the article was read, the *Ss* filled out a questionnaire in which they were asked to indicate their judgments of fairness of presentation and propagandistic intent of the author, their own opinions on the topic of the article, their knowledge of the content of the article as tested by a set of multiple-choice items, and finally their opinions on the topic of the other article and on five issues not dealt with by either article. The before-opinion items on the topics covered by the communications were repeated in the after-questionnaire.

Following this, the *E* asked the *S* to express his judgment of the worthwhileness of the tests, whether there was anything he particularly liked or disliked about the experimental session, and if he would be interested in returning for another session at a later time. The first *E* returned then and informed the *S* of the nature of the experimental situation. Several interview questions were asked concerning *S*'s reactions to his "failures" on the first two tests and to the disparaging remarks made by *E*, his ability to concentrate adequately on the article, and the possibility that he thought about *Es*' opinions on the topic of the article and the influence such thoughts might have had on him.

Scoring. The scoring of the Rosenzweig P-F Study and the opinion responses was the same as that employed by Weiss and Fine (14). Each picture of the former was scored zero or one depending on the presence or absence of an extra-punitive response; thus, the theoretical range was zero to 24. The opinion responses were scored -1 for an expression of leniency, zero for uncertainty, and $+1$ for punitiveness; the theoretical range was -4 to $+4$. Each TAT story was assessed on a five-point basis, zero to four, with zero being assigned for absence of aggression and four for violent aggression.

RESULTS

Before examining the effects of the communications on opinions, it is necessary to determine if the experimental procedures were successful in producing a separation of the groups on the dimension of aggression arousal. The interview material indicated that two-thirds of the *Ss* subjected to the aggression-arousal treatment reported experiencing feelings of hostility toward *E*, and almost all felt frustrated in their attempts to perform well.

[2] Weiss acted as the first *E* for all *Ss* and Fine as the second *E*. Only one *E* was in the room with the *S* at any time.

TABLE 1

MEANS FOR TAT STORIES OF EXPERIMENTAL GROUPS

Card	Aggressive	Nonaggressive	p(one tail)
18GF	2.20	2.24	—
3BM	1.81	1.35	.06
Total	4.02	3.60	>.15

TABLE 2

MEAN BEFORE-OPINIONS AND CHANGES IN OPINIONS ON TOPICS OF ARTICLES (BASED ON 4 ITEMS)

Opinions on Delinquency

Experimental Group		Before-Opinion	Change Score*
Read article	Aggressive	−2.88	+1.83
	Nonaggressive	−2.25	+0.38
Did not read article	Aggressive	−2.04	−0.21
	Nonaggressive	−2.88	−0.04

Opinions on America's Relation with Allies

Experimental Group		Before-Opinion	Change Score*
Read article	Aggressive	−0.04	−0.54
	Nonaggressive	−0.46	−1.29
Did not read article	Aggressive	+0.42	+0.04
	Nonaggressive	−0.38	+0.08

* A plus sign signifies change toward punitiveness; a minus sign indicates change toward leniency.

The protocols frequently contain such comments as "very angry," "burnt up," "wanted to tell you to shut up," "wondered what value college was going to be to me," "very upset at poor performance," "never experienced such ego deflation so quickly." Only *S*s from this group said they would not volunteer for another session, more of these questioned the value of the tests than did those of the other group, and only members of this group said they disliked features of the general experimental situation. The significant difference between the groups on the second TAT story, as indicated in Table 1, supports the interview results.[3] Many of the aggression-aroused *S*s said that they viewed the TAT as a second chance to show creativity and were hopeful of improving their performances. But *E*'s derogatory comments after the first story were a severe setback and aggravated feelings of anger and hostility. Most of them still exhibited hostility and concern when the first *E* returned to the room and expressed great relief at being informed of the nature of the experiment. On the contrary, the other group of *S*s reported pleased, satisfied sentiments with their performances; such comments as "felt pleased with myself," "very happy I did well," "good to know you have some creative imagination" were frequently made. It seems evident then that the desired effects of the experimental treatments were adequately realized.

Before proceeding to the main findings of the research, two other items of relevant information should be mentioned. None of the *S*s in either group reported in the interviews that they had been concerned at all with assumptions about *E*'s opinion on the topic of the article read, either while reading the article or answering the questionnaire. Also, the distributions of extrapunitive scores, based on the Rosenzweig measure, for the four treatment groups were not reliably different from each other.

Table 2 indicates the effects of the articles on opinions. There were no reliable differences among the groups in before-opinions on either issue. On the topic of delinquency, all groups were decidedly on the lenient side of the attitude dimension; the means for the other topic however were close to zero.[4] Both articles affected opinions significantly on their respective issues. The difference between the aggressive and the nonaggressive groups who read the article on delinquency is significant at the .02 level (one tail); and for the other topic, the difference has a p value of .08 (one tail.) The induced affects alone did not produce reliably different changes in opinions on the topics, if the relevant article had not been read.[5]

[3] In all the tables, comparisons between the total aggressive and nonaggressive groups are based on *N*'s of 48 for each. Comparisons between treatment-article subgroups are based on *N*'s of 24 for each.

[4] A small, pre-experimental sample of *S*s provided a somewhat punitive opinion mean on policy toward America's allies. If it had been known that the mean for the over-all experimental groups would be close to zero, this topic alone could have been used for all treatments. One communication would have urged leniency and the other punitiveness. Methodologically, this would have been a more precise procedure. However, additional control groups would have been needed to determine the effect on opinions of the experimental inductions alone.

[5] The second-order difference on the topic of delinquency is significant at the .03 level, one tail; and

TABLE 3

MEAN SCORES ON FACT-QUIZ TESTS

Group	Article	
	Delinquency (6 items)	Allies (5 items)
Aggressive	4.21	4.42
Nonaggressive	4.58	4.54
p value (two tails)	$>.30$	$>.60$

TABLE 4

PERCENTAGE OF EACH GROUP JUDGING ARTICLE READ AS "FAIR" OR "PROPAGANDA"

Group	Article	"Completely" or "Mostly" Fair	"Propaganda"
Aggressive	Allies	83.3	29.2
Nonaggressive	Allies	83.3	12.5
Aggressive	Delinquency	29.2	45.8
Nonaggressive	Delinquency	33.3	37.5

The possibility was considered of a relation between reading-responding to one of the communications and the expression of lenient-punitive opinions on topics not the focus of the article read. With the data at hand, this was examined in two ways: by testing for an association between changes in opinions on the topics of both articles and for one between changes in opinions on the topic of the article read and the five after-items dealing with lenient-punitive policies on topics not the focus of either article. Neither relationship was significant. Furthermore, the aggressive and nonaggressive groups did not differ significantly in their responses to the five special after-items.

Acquisition of the content of the communications was examined by a set of multiple-choice items containing four alternative responses for each. The data in Table 3 reveal that, as measured, there were no reliable differences between the relevant groups in learning the communication content.

Two questions eliciting *S*s' judgments of the possible propagandistic intent and fairness of presentation of the communications were included in the after-questionnaires. The obtained data are presented in Table 4. None of the differences between the groups reading the same article is significant. Also, the article on delinquency is less frequently judged as "fair" and more frequently judged as "a piece of propaganda" than is the article on the relations of our allies with communist China. A separate analysis of the relation between judgments and opinion change indicated that those judging the delinquency article as "fair" were reliably more influenced than were those judging it as "one-sided"; judgments of this article on the dimension of propagandistic intent were not however related reliably to communication effectiveness. The effect of judgments on the opinion influence of the other article could not be adequately analyzed on account of the small numbers of *S*s considering it as "propaganda" or "one-sided."

DISCUSSION

The data support the main hypotheses that aggression-aroused *S*s will be more influenced by a punitively oriented communication and less influenced by a leniency oriented one than will nonaggressive *S*s. The smaller differential effect of the article on America's allies may be due to the relatively nonspecific nature of the persuasion; that is, the expressed viewpoint argued more for viewing the problem of relations with China through the eyes of other nations than for specific policy recommendations, which represented the focus of the opinion items. Previous research indicates that opinions which are the focus of specific communication appeals are influenced more than opinions that are indirectly related to the appeals (10). Also, the possible difference in the changeability of opinions on the two topics should be considered.

The term nonaggressive rather than ego-satisfied has been used to designate one group of *S*s, despite evidence from the interview material that some degree of ego satisfaction was achieved. To maintain a theoretical and experimental distinction between nonaggressiveness and ego satisfaction requires evidence that they differentially affect communication acceptance. The nonaggressive control group needed to determine this was unfortunately not included in the study.

the difference on the other topic does not approach significance. These differences are obtained by reducing the difference between the aggressive and nonaggressive groups that read the article by the difference between the groups not reading the article.

The results are consonant with a frustration-aggression-displacement interpretation. The "failure" technique obviated achievement of a desired goal, which reflected strongly on the self-image and self-esteem of the person (11). The accompanying insults facilitated the arousal of an instigation to external aggression, which could not be expressed against the higher-status and more "powerful" instigator. This response predisposition aided the acceptance of a punitively oriented communication, which functioned as an outlet for the expression of aggression with little fear of punishment. Whether this displacement served to reduce the facilitating predisposition to be influenced by following aggressively oriented communications is not known. Previous theory (5) and research (7, 13) suggest the possibility of such a consequence. If this did occur, a socially relevant, cathartic use of mass media can be conceived. Another implication is that the effectiveness of communications ordered on a leniency-aggressiveness dimension may be influenced by the immediately prior and consequent communication context (8).

The incongruency between the proffered response "bait" and the response predisposition of the aggressively aroused *S*s is suggested as the basis for the results on the issue of America's allies. This interpretation as well as that for the effects of the delinquency article may require supplementation if it is shown that ego-satisfaction engenders a similar but opposite predisposition to communication acceptance as that provoked by aggression arousal.

Some analysts of the genesis and maintenance of social prejudices have adopted in part an approach based on a frustration-aggression thesis (3, 4, 5, 6, 9). The current findings are clearly relevant to such interpretations. However, the over-all influence of the communications regardless of experimental inductions should be kept in mind as a cautionary note. The presence of individual differences in response to the treatment-communication combinations must also be considered; with a larger sample, the extrapunitive score could have been used to produce adequate subgroups for reliable examination of the contribution of this response tendency to individual differences. Finally, the evidence is not available in this study for determining the effect of prior reduction in aggressive needs on response to communications.

Aggression arousal was not sufficient by itself to effect changes in opinions. This finding is not consonant with the data of Miller and Bugelski (12). The different response methodology which they employed may be a factor contributing to the discrepant results. In their study, an increase in displaced aggression following a frustrating experience was evidenced by the assignment of favorable and unfavorable traits to certain nationality groups. This technique may be a more sensitive method than the elicitation of specific opinion responses on policy. Also, opinions on issues which have been in the focus of public attention or to which some prior thought has been given may be less susceptible to fluctuation without some communication pressure than those on topics without such characteristics. Given this qualification, our position is that aggression arousal predisposes a person to certain opinion changes; but for the predisposition to achieve significant specific effects, communication appeal is needed.

There is some evidence in both this research and the previous one by the authors (14) indicating that judgment of fairness of presentation is more closely related to communication effectiveness than is judgment of propagandistic intent.[6] Some of the spontaneous comments of a number of *S*s suggest an explanation for this. Propagandistic intent was equated by them with motivated persuasion for a particular position. The latter was attributed to all communicators; that is, everyone has an "ax to grind." Thus, this judgment was not as discriminating with respect to the persuasiveness of a communication. This finding, if supported by other studies and analyses, has evident implications for communication effectiveness and resistance.

SUMMARY

The research was designed to test the hypotheses that aggressively aroused persons are influenced more by a punitively oriented communication and less by a leniency-oriented one than are nonaggressive persons. A before-after design was employed. One main group was exposed to a failure-insult pro-

[6] A positive but not strong association between the two exists.

cedure to evoke an instigation to aggression and the other to an ego-satisfying experience. Then, one half of each group read a communication urging punitiveness toward delinquents and the other half read a communication suggesting leniency and consideration in America's relations with her allies. An immediate-after opinion questionnaire followed. After being informed of the nature of the experiment, the *S*s were interviewed to determine their reactions to the procedures.

The data support the research hypotheses. Also, it was found that an instigation to aggression alone does not seem to affect opinions on leniency-punitive issues which are not dealt with by the communication read. The implications of the results for theory and further research were noted.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22,** 9–39.
3. ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass.: Addison-Wesley, 1954.
4. BETTELHEIM, B., & JANOWITZ, M. *Dynamics of prejudice.* New York: Harper, 1950.
5. DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.
6. FENICHEL, O. Elements of a psychoanalytic theory of anti-semitism. In E. Simmel (Ed.), *Anti-semitism: a social disease.* New York: International Univer. Press, 1946. Pp. 11–32.
7. FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50,** 3–11.
8. FLOWERMAN, S. H. Mass propaganda in the war against bigotry. *J. abnorm. soc. Psychol.*, 1947, **42,** 429–439.
9. HOROWITZ, E. L. Race attitudes. In O. Klineberg (Ed.), *Characteristics of the American Negro.* New York: Harper, 1944. Pp. 139–252.
10. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communications.* Princeton: Princeton Univer. Press, 1949.
11. MASLOW, A. H. Deprivation, threat and frustration. *Psychol. Rev.*, 1941, **48,** 337–352.
12. MILLER, N. E., & BUGELSKI, R. Minor studies in aggression: the influence of frustration imposed by the in-group on attitudes expressed toward out-groups. *J. Psychol.*, 1948, **25,** 437–442.
13. THIBAUT, J. W., & COULES, J. The role of communication in the reduction of interpersonal hostility. *J. abnorm. soc. Psychol.*, 1952, **47,** 770–777.
14. WEISS, W., & FINE, B. J. Opinion change as a function of some intrapersonal attributes of the communicatees. *J. abnorm. soc. Psychol.*, 1955, **51,** 246–253.

Received February 9, 1955.

THE INFLUENCE OF REINFORCEMENT AND PUNISHMENT IN A MINIMAL SOCIAL SITUATION[1]

JOSEPH B. SIDOWSKI, L. BENJAMIN WYCKOFF, AND LEON TABORY
University of Wisconsin

THE study of conditioning in the laboratory, and the development of conditioning theory (3, 5, 6), are justified in part by the hope that the principles derived will ultimately enhance our understanding of human behavior in complex situations including social ones. The realization of such hopes is commonly relegated to some future time, and the path along which we expect to proceed is specified in the vaguest terms. In attempting to approach this problem, several questions may be raised: (*a*) Is it possible to handle social behavior within the framework of current conditioning theory, and if not, at what point along a scale of increasing complexity is it necessary to introduce additional concepts? (*b*) What new conceptual and analytical tools are necessary to handle the additional complexity encountered?

In the present study we attempt to reevaluate the essential features of a social situation as viewed entirely within the framework of conditioning theory, and to investigate the simplest situation that could be considered truly social within this framework. In this way, we hope to lay the groundwork for a systematic investigation of social situations of increasing complexity, and to clarify methodological and theoretical problems. Analysis of social behavior in these terms implies that no assumption whatever is to be made regarding such concepts as attitude, understanding, awareness, etc. If such factors enter the picture at all, they are to be considered as manifestations of complex habits. Preferably, they will not be encountered at all in the minimal social situation, and in any event cannot be part of the definition of social behavior.

When restrictions such as these are accepted, we arrive at a conception of social behavior that will, no doubt, seem strange to most social psychologists. However, we gain the advantage of having a system that allows for a high level of objectivity in our definitions and principles, and one that is readily related to the mass of careful investigations of conditioning (3, 4, 5, 6). In this analysis we will assume that the main factors controlling social behavior are reinforcement (reward) and punishment. Within this framework, the essential features of a social situation may be expressed as follows: (*a*) Two or more *S*s have at their disposal responses which result in reinforcing or punishing effects on other *S*s. (*b*) The principal sources of reinforcement and punishment for any *S* depend on responses made by other *S*s. (*c*) The responses controlling reinforcement and punishment are subject to learning through trial and error. We will assume that, whatever else may be involved, at least the above features are present in any social situation. At this point we may ask what learning would be predicted on the basis of these conditions alone, assuming only that the principles of conditioning theory are operating. In the present experiment an attempt was made to place *S*s in a situation involving only these minimal features.

Two *S*s isolated from each other were provided with push buttons by means of which each could give the other reinforcement or punishment (score or shock). The effects of previous social learning were minimized, since *S*s were not told (nor did they guess) that they were in a social situation at all.

In this situation, if *S* makes a response, he does not "suffer the consequences" directly. An *S* can influence the reinforcing and punishing features of his environment, only to the extent that his actions may change the behavior of the other *S*. Thus, if *A* makes a response which results in a reinforcement for *B*, we may expect *B* to continue with, or to repeat, his immediately preceding behavior. Similarly, if *A* makes a response which results in a punishment for *B*, we may expect *B* to tend to discontinue his immediately preceding activity. In addition, when *B* receives punishment, we will also expect an increase in his over-all activity level. These changes in *B*'s behavior have a direct effect on the reinforcement and punishment being received by *A*.

[1] The writer wishes to thank Dr. David A. Grant for his helpful suggestions.

Thus *A* does receive some consequences of his responses through an indirect but not entirely unlawful process.

It is not intended that this situation should be directly comparable to familiar everyday encounters. We generally enter social situations already equipped with a massive amount of prior learning relating to the effects of our behavior on others. For example, we have stylized ways of indicating approval or disapproval that are in effect methods of reinforcing or punishing others. On the other hand, in many situations, the responses we have learned are no longer appropriate. For example, an adult who has had little experience with children may find that his habitual ways of expressing approval and disapproval are completely ineffective when he first begins to interact with a child. As he gains experience he learns new responses. Other illustrations could be cited. One occasionally encounters adults who react negatively to conventional reinforcements or punishments. The process through which this learning and relearning takes place will be central to any extension of conditioning theory into the realm of social behavior. We may ask, is face-to-face contact and "understanding" of the relationship between *S*s necessary for this learning or would such learning occur simply as a result of the operation of reinforcement and punishment? It is questions such as these which we hope to answer through the present research program. The first step in this program is an investigation of behavior in the minimal situation.

The experiment. As indicated above, two *S*s, isolated from each other, were given the opportunity to make responses which resulted in reinforcement or punishment (score or shock) for the other *S*. The general objective of the experiment was to discover what the course of learning would be under these conditions. It seemed likely that one of the most important parameters which would affect performance would be the magnitude of punishment used. Thus, two widely separated levels of shock were used for different groups of *S*s.

METHOD

Subjects. The *S*s were 40 members of elementary psychology courses at the University of Wisconsin who received class points for participating.

Apparatus. The two experimental rooms were each equipped with a control panel which provided two push buttons, an electrically operated counter, and a pair of electrodes. The right and left buttons in Room 1 controlled the counter and shock circuit in Room 2, and vice versa. Thus, pressing the right-hand button in Room 1 registered one point on the counter in Room 2, while pressing the left-hand button in Room 1 delivered an electric shock to the electrodes in Room 2. Similarly the buttons in Room 2 controlled the counter and shock in Room 1, except that the right and left buttons had the opposite effect. A small red light also blinked each time the counter was activated, and a white light served as a signal that the experiment had begun. The white light remained on until the session ended. The *E*'s control board contained counters which indicated the number of times that each of the four buttons had been pressed.

An electronic circuit provided for regulation of the shock current and automatic compensation for changes in *S*'s resistance. Separate controls were available for each of the two *S*s.

Procedure. The 40 *S*s, 20 pairs, were randomly assigned to two groups of 10 pairs each. The Strong-Shock group received shock current equal to 200 per cent of their absolute threshold, while the Weak-Shock group received 110 per cent of the threshold reading. Preliminary tests indicated that the absolute threshold was more reliable and more clearly defined than a "pain" threshold, so this measure was used as a basis for shock levels. The current used for the Strong-Shock group had a mean of .12 ma. and a standard deviation of .05, while the current used for the Weak-Shock group had a mean of .12 ma. and an *SD* of .07.

Although two *S*s served during each experimental session, each *S* was led to believe that he was the only *S* serving in the experiment. At no time during or after the session were the *S*s given any indication that another *S* had served in the experiment. Any questions concerning the current were ignored; however, each *S* was offered the opportunity to withdraw from the experiment before the session began. None of the *S*s withdrew.

After the threshold readings were completed, *S* was taken to one of the two experimental rooms where he was seated before the *S* control board. Electrodes were attached to the first two fingers of his left hand. Instructions were then read to *S*, after which he was told to relax for a few minutes while the apparatus warmed up. During this time, *E* brought the second *S* before the *E* panel and went through the same procedure. The second *S* was taken to Room 2 and seated before the second control board. All *S*s were given the following instructions:

"When the white light goes on the experiment begins. The experiment is in progress as long as the white light stays on. You can press either of these buttons in any manner that you wish, and as frequently as you wish. Do not attempt to press both buttons at the same time and use only your right hand for pressing. You are to keep your left hand on the table. Do not hold a button down, but just press and release it. The object of the experiment is to make as many points as you can. Your point score will appear on this counter. The red light will blink and the counter will turn each time that you score a point. When the white light goes off, it is a signal that the

experiment has ended. Remain seated until *E* comes into the room. Remember, you are to try to make as high a score as you can in any manner that you can."

During the experiment, the four scores, a shock and a point score for each of the two *S*s, were recorded by *E* every 30 sec. over a period of 25 continuous minutes. This gave a record of 50 scores for each button on the two *S* control boards.

At the conclusion of the experimental session, a series of questions was asked in order to determine whether or not any of the *S*s were aware of the presence of another *S* in the experiment.

RESULTS

In the initial analysis, scores over 5 minute intervals were tabulated for "score" and "shock" responses for each of the two groups. These data are presented graphically in Fig. 1. An Alexander trend test (1) was carried out on the data and the *F* was found to be significant beyond the .05 level of confidence. Duncan's test (2) for differences between ranked treatments was carried out on the ranked means (presented in Table 1) of the four button responses. The differences between all ranked means were significant well beyond the .05 level of confidence (an exception was the nonsignificant difference between the Weak-Shock—Weak-score means). In addition a ratio for each 5-minute interval was computed to give the proportion of the total responses which were "score" responses. These data are presented in Fig. 2. An Alexander trend test showed an *F* significant beyond the .05 level of confidence.

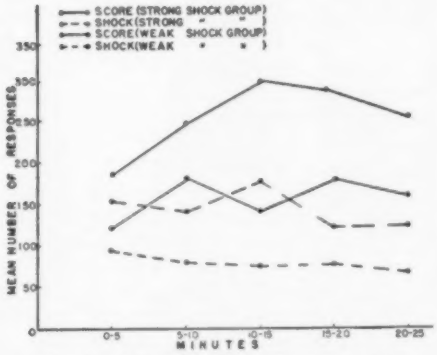


FIG. 1. MEAN NUMBER OF RESPONSES ON THE SHOCK AND SCORE BUTTONS BY THE STRONG-SHOCK AND WEAK-SHOCK GROUPS

TABLE 1

RANKED MEANS FOR THE FOUR-BUTTON RESPONSES

Ranks	Means
Strong-shock button	397.85
Weak-shock button	719.95
Weak score button	786.75
Strong score button	1,288.65

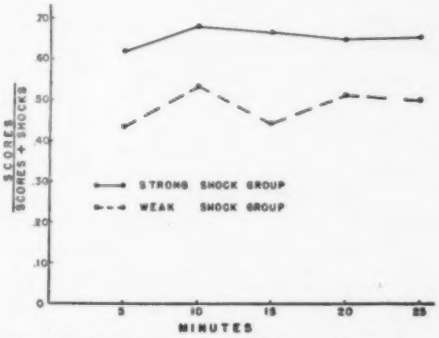


FIG. 2. PROPORTION OF THE TOTAL NUMBER OF RESPONSES WHICH WERE CORRECT RESPONSES (SCORE RESPONSES) DURING THE 25-MIN. EXPERIMENTAL SESSION

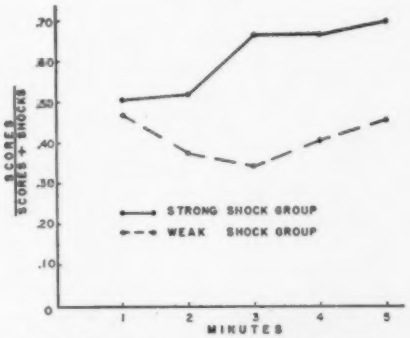


FIG. 3. PROPORTION OF THE TOTAL NUMBER OF RESPONSES WHICH WERE CORRECT (SCORE RESPONSES) DURING THE FIRST 5 MINUTES

These tests showed that the Strong-Shock group yielded significantly more "score" than "shock" responses, while no such difference for the Weak-Shock group was found. The tests did not indicate that the slope of the curves, which would have reflected gradual improvement, was significant. Inspection of the curves suggested that our failure to show significant slopes might have been due to the fact that much of the learning had occurred within the

first five minutes. Therefore, the data for the first five minutes were divided into 1-min. intervals and analyzed. These data are presented in Fig. 3. An Alexander trend analysis showed a significant difference between group slopes beyond the .05 level of confidence. This confirms our observation that learning took place within the first five minutes.

As a result of the preceding analyses, the major fact to be noted is that learning did occur in the Strong-Shock group, whereas no evidence of learning was obtained in the Weak-Shock group.

Responses to the questions at the end of the experiment indicated that none of the *S*s were aware that a second *S* was involved.

DISCUSSION

Two major questions may be raised regarding the experimental findings: (*a*) Why did learning appear in the Strong-Shock group but not in the Weak-Shock group? (*b*) Why did learning occur at all under the conditions of this experiment?

In order to answer these two questions it is necessary to examine some of the sequences of responses which may have occurred during the course of the experiment. Examination of all possible sequences of events that might occur in the two groups becomes too cumbersome for our present purposes. However, it will be of interest to consider a few examples of possible beginning sequences and possible end states. Generally, we might assume random behavior by both *S*s at the beginning of the experimental session. For example, let us suppose that each *S* made a response that shocked the other. It is not unlikely that this event might occur by accident early in the session. However, it is clear that some shift away from this behavior would soon appear since neither *S* is receiving reinforcement and both are being punished. This shift would probably occur sooner under conditions of Strong Shock. The *S*s might shift at the same time, or one might shift sooner than the other, resulting in at least two possible new patterns. If both shifted to the reinforcement button at the same time, both would now receive reinforcement and we would expect a tendency for this pattern to continue. (In fact, the same prediction would be made if *S*s commenced to reinforce each other at any time.) This pattern is perhaps the most obvious "self-sustaining" pattern which might emerge, and thus is one of the possible end states which might prevail. On the other hand, suppose that one *S* shifted to the reinforcement button before the other, or that, at any time, one *S* pressed the reinforcement button while the other pressed the shock button. The *S* who is pressing the reinforcement button should tend to shift to the shock button, with the result of reversion to the condition of both shocking each other. Since both of these patterns are "unstable," we might obtain a continual oscillation between the two states, with one *S* and then the other shifting temporarily to the reinforcement button. This oscillation itself would form a new pattern which could be stable under certain conditions, particularly if the magnitude of the shock were relatively small. If this magnitude were small enough, the intermittent reinforcement obtained by both *S*s might sustain the over-all pattern. This oscillatory pattern thus represents another possible end state, one that would be most likely to occur under conditions of Weak Shock. Under conditions of Strong Shock, we might expect that the intermittent reinforcement obtained by both *S*s would not sustain the over-all pattern of oscillation because of the punishing effects of the shock. Of course, many other sequences of responses could occur during the session, but it appears that most of these would be unstable.

Although the above considerations are not complete, they suggest a possible explanation for the occurrence of learning in the present experiment and the differential effect of the two levels of shock. It is suggested that under conditions of Strong Shock, the most probable stable pattern (end state) was that of both *S*s reinforcing each other. Thus, *S*s would tend to shift among patterns until they hit on this solution. (Some *S*s could be expected to return to earlier responses, such as pressing the shock button from time to time due to the intermittent reinforcement which may have occurred early in the session following this type of response.) Under conditions of Weak Shock, the pattern of oscillation between the buttons was presumably the most probable stable state since any of several initial sequences could lead into this pattern. It was not possible to

analyze the sequential patterns in the present experiment, so further experimentation would be necessary to test these interpretations.

Within the framework of the present experiment, we can conclude that learning was demonstrated in a minimal social situation where: (*a*) the effects of previous social learning had been minimized, and (*b*) reinforcement and punishment were the only observable factors influencing the behavior of the *S*s.

SUMMARY

The present experiment investigates the influence of reinforcement and punishment in a minimal social situation. Two *S*s, isolated from each other, were provided with two push buttons by means of which each could give the other a shock or score. The effects of previous social learning were minimized since *S*s were unaware of the fact that they were in a social situation at all.

The *S*s were divided into a Weak Shock and a Strong Shock group. All *S*s were merely told that they could press the two buttons in any manner that they pleased (no information was given concerning the purpose of the buttons). Each *S* was also told that he was to make as many points as possible. His score was indicated on a counter which was mounted in front of *S*.

It was found that the proportion of the total number of responses (both score and shock) that were correct (score responses) was significantly higher for the Strong-Shock group. Learning occurred in the Strong-Shock group within the first 5 minutes of the 25-minute experimental session. Learning was not, however, evident in the Weak-Shock group.

Principles of reinforcement were used to interpret the results.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. DUNCAN, D. B. A significant test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2**, 171–189.
3. KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.
4. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
5. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
6. SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

Received February 14, 1955.

THE EXPERIMENTAL CHANGE OF A MAJOR ORGANIZATIONAL VARIABLE[1]

NANCY C. MORSE
The Merrill-Palmer School

AND

EVERETT REIMER
Office of Personnel, Commonwealth of Puerto Rico

THIS experiment is one in a series of studies of social behavior in large-scale organizations undertaken by the Human Relations Program of the Survey Research Center. Its primary aim is to investigate the relationship between the allocation of decision-making processes in a large hierarchical organization and (*a*) the individual satisfactions of the members of the organization, (*b*) the productivity of the organization.

The results of several previous studies suggested that the individual's role in decision-making might affect his satisfaction and productivity. The effectiveness of decision-making in small groups shown by Lewin, Lippitt, and others (4, 5) and the successful application of small-group decision-making to method changes in an industrial setting by Coch and French (1) both indicated the possibilities for enlarging the role of the rank and file in the ongoing decision-making of an organization. The practical experience of Sears, Roebuck and Co. with a "flat," administratively decentralized structure, described by Worthy (8), pointed in the same direction, as did the survey findings by Katz, Maccoby, and Morse (2) that supervisors delegating greater authority had more productive work groups. The logical next step seemed to be the controlled testing of hypotheses concerning the relationship between role in organizational decision-making and two aspects of organizational effectiveness: satisfaction and productivity. Two broad hypotheses were formulated:

Hypothesis I. An increased role in the decision-making processes for rank-and-file groups increases their satisfaction (while a decreased role in decision-making reduces satisfaction).

Hypothesis II. An increased role in decision-making for rank-and-file groups increases their productivity (while a decreased role in decision-making decreases productivity).

Both these hypotheses deal with the effects on the rank and file of different hierarchical allocations of the decision-making processes of the organization. The rationale for the satisfaction hypothesis (I) predicts different and more need-satisfying decisions when the rank and file has decision-making power than when the upper echelons of the hierarchy have that power. Furthermore, the process of decision-making itself is expected to be satisfying to the majority of people brought up in American traditions. Underlying the productivity hypothesis (II) was the consideration that local unit policy-making would increase motivation to produce and thus productivity. Motivation should rise when productivity becomes a path for greater need satisfaction. The productivity hypothesis predicts a higher degree of need satisfaction (as does Hypothesis I) *and* an increase in the degree of dependence of satisfactions upon productivity under conditions of greater rank-and-file decision-making. It is expected that when rank-and-file members work out and put into effect their own rules and regulations, their maintenance in the organization (and thus their satisfactions) will depend much more directly upon their performance.

[1] This is a short description of an experiment done while the authors were on the staff of the Human Relations Program of the Survey Research Center, University of Michigan. Financial support for field work and analysis of the data came from the Rockefeller Foundation, the Office of Naval Research Contract No. N6 onr-232 Task Order II, and the company in which the research was done. In addition to the authors the field staff of the experiment included: Arnold Tannenbaum, Frances Fielder, Gilbert David, Arlene Kohn Gilbert, Barbara Snell Dohrenwend, Ann Seidman, Jean Kraus Davison, and Winifred Libbon. The analysis staff included: Nancy Morse, Arnold Tannenbaum, Arlene Kohn Gilbert, and Ruth Griggs. The experiment will be described fully in a book now in preparation. Floyd H. Allport provided extensive assistance on the theoretical problems of the study. The experiment was under the general direction of Daniel Katz, director of the Human Relations Program during the field phase of the experiment, and Robert Kahn, director of the Human Relations Program during the analysis phase. The authors wish to express their appreciation to the staff members on the experiment and to the people in the company who cooperated in the experiment. They also want to thank particularly Daniel Katz, Robert Kahn, Arnold Tannenbaum, Carol Kaye, and Jane Williams for their helpful comments and criticisms of this article.

PROCEDURE

The experiment was conducted in one department of a nonunionized industrial organization which had four parallel divisions engaged in relatively routine clerical work. The design involved increasing rank-and-file decision-making in two of the divisions and increasing upper-level decision-making in the other two divisions. The time span was one and one-half years: a before measurement, one-half year of training of supervisors to create the experimental conditions, one year under the experimental conditions, and then remeasurement. The two pairs of two divisions each were comparable on relevant variables such as initial allocation of the decision-making processes, satisfaction and productivity, as well as on such background factors as type of work, type of personnel, and type of supervisory structure.

The rank-and-file employees were women, mostly young and unmarried, with high school education. The usual clerk's plans were for marriage and a family rather than a career. The population used in the analysis except where noted is a subgroup of the clerks, the "matched" population. These clerks were present throughout the one and one-half year period, and their before and after questionnaires were individually matched. While they comprise somewhat less than half of the clerks present in these divisions at any one time, they are comparable to the total group, except on such expected variables as length of time in the division, in the work section, and on the job.

One aspect of the work situation should be mentioned, as it bears on the adequacy of the setting for a test of the productivity hypothesis. The amount of work done by the divisions was completely dependent upon the flow of work to them, i.e., the total number of units to be done was not within the control of the divisions. With volume fixed, productivity depends upon the number of clerks needed to do the work, and increased productivity can be achieved only by out-placement of clerks or by foregoing replacement of clerks who leave for other reasons.

The Development of the Experimental Conditions

Creating the experimental programs included three steps: (*a*) planning by research staff and company officials; (*b*) introducing the programs to the division supervisory personnel and training of the supervisors for their new roles; and (*c*) introduction to the clerks and operation under the experimental conditions.

The experiment was carried out within the larger framework of company operations. The introduction, training, and operations were in the hands of company personnel. The experimental changes were not made through personnel shifts; the changes were in what people did in their jobs with respect to the decision-making processes of the organization.

Two main change processes were used in both the Autonomy program, designed to increase rank-and-file decision-making, and in the Hierarchically-controlled program, designed to increase the upper management role in the decision-making processes. First, there were formal structural changes to create a new organizational environment for the divisions in each program. In both programs the hierarchical legitimization of new roles preceded the taking of the new roles.[2] In the Autonomy program authority was delegated by upper management to lower levels in the hierarchy with the understanding that they would redelegate it to the clerical work groups. In the Hierarchically-controlled program, authority was given to the higher line officials to increase their role in the running of the divisions and to the staff officials to increase their power to institute changes within the two divisions in that program. Second, there were training programs for the supervisors of the divisions to ensure that the formal changes would result in actual changes in relations between people. (For a longer description of the change programs see Reimer [6].)

Measurement

The results of the changes were gauged through before and after measurements and through continuing measurements during the experimental period. The major emphasis was on the attitudes and perceptions of the clerks as reflected in extensive questionnaires. In addition, the training programs and the operations phase of the experiment were observed. Before and after interviews were conducted with the supervisory personnel of the division. Data from company records such as productivity rates, turnover figures, etc., were also included.

The data reported here will be confined to material most pertinent to the testing of the two hypotheses. For other related aspects of the experiment, see Tannenbaum's study of the relationship of personality characteristics and adjustment to the two programs (7), Kaye's study of organizational goal achievement under the Autonomy program (3), as well as forthcoming publications.

RESULTS[3]

Success of Experimental Manipulation

The first question was to discover whether or not the change programs were successful in creating the conditions under which the hypotheses could be tested. Two types of data are pertinent. The first is descriptive data concerning the actual operations of the two programs. The second is perceptual data from the clerical employees themselves indicating the degree to which they saw changes in their role in organizational decisions.

[2] Weber and others have used the word "legitimization" to refer to the acceptance by subordinates of the authority of superiors. We are using the word in quite a different sense. By hierarchical legitimization we mean the formal delegation of authority by superiors to subordinates. This delegation *legitimizes* the subordinates' utilization of this authority.

[3] For the statistical tests used in this section, we have assumed that the individuals were randomly chosen, while the selection of individuals by divisions undoubtedly results in some clustering effect. The levels of significance should, therefore, be considered as general guides rather than in any absolute sense.

The operations of the divisions in fact changed in the direction expected. In the Autonomy program the clerical work groups came to make group decisions about many of the things which affected them and which were important to them. The range of the decisions was very great, including work methods and processes, and personnel matters, such as recess periods, the handling of tardiness, etc. Probably the most important area in which the clerks were not able to make decisions was the area of salary. Some of the work groups were more active in the decision-making process than others, but all made a very great variety of decisions in areas important to them. In the Hierarchically-controlled program the changes decreased the degree to which the employees could control and regulate their own activities. One of the main ways in which this greater limitation was manifested was through the individual work standards that staff officials developed for the various jobs. Also the greater role of upper line and staff officials in the operation of the divisions meant that the indirect influence which the clerks could have on decisions when they were made by division managers and section supervisors was reduced.

The clerks were operating under different conditions in the two programs as the result of the experimental changes, but did they perceive these changes? The method of measuring the perception of changes in decision-making was by asking clerks about their part and about the part of people above their rank in decisions with respect to a wide variety of areas of company operations, or company systems. The following questions were asked about each major area of company operations or system: "To what degree do company officers or any employees of a higher rank than yours decide how the _____System is set up and decide the policies, rules, procedures or methods of the _____System?" (followed by a line with the landmark statements: not at all, to a slight degree, to some degree, to a fairly high degree and to a very high degree) and, "To what degree do you and the girls in your section decide how the _____System is set up and decide the policies, rules, procedures or methods of the _____System?" (followed by a line with the same landmark statements as the first question).

The extreme degree of perceived hierarchical control of the decision-making processes would be shown by the clerks answering that employees of a higher rank than theirs made the decisions, "to a very high degree" and the clerks made them "not at all." Table 1 shows the number of systems where there are half or more of the clerks endorsing these two statements for the before situation and for the two experimental situations. (The Autonomy program is designated in Table 1 and thereafter as Program I and the Hierarchically-controlled program as Program II.) Questions were asked for 27 company systems in the before measurement and 24 systems in the after measurement.

Table 1 shows that the clerks perceived the decision-making processes for most of the company operations measured as located at hierarchical levels above their own, prior to the introduction of the experimental changes. The experimental changes in the Autonomy program divisions resulted in their seeing decision-making activities as much less exclusively confined to levels above theirs. The changes in the Hierarchically-controlled program were less striking but they resulted in the clerks judging that all of the systems about which they were asked in the after situation had their policies molded to a very high degree by people above their level.

The relative role of the hierarchy compared to the rank and file as perceived by the clerks was measured by assigning scores from 1 to 9 for the landmark positions on the scales for the two questions and then dividing the score for upper-level decision-making by the score for rank-and-file decision-making. The theoretical

TABLE 1

NUMBER OF COMPANY SYSTEMS IN WHICH CLERKS PERCEIVE VERY HIGH UPPER LEVEL CONTROL OF DECISION-MAKING ALLOCATION

Response	Number of Systems in which Half or More Clerks gave Specified Response		
	Before All Divs	After	
		Program I Divs	Program II Divs
Upper levels decided policies to a very high degree	20	7	24
Clerks did not decide policies at all	25	9	23
Total number of systems measured	27	24	24

range for the resulting index is from 9.0 to 0, with numbers less than 1 indicating greater local control than upper-level control. Table 2 includes the average index scores for the systems from the before and after measurements calculated by division.

Table 2 indicates the change in the divisions in the Autonomy program toward greater perceived rank-and-file role in decision making, but also shows that the upper levels are seen as still having the major role in the after situation. (The downward shift in perceived decision-making control in the Autonomy program is significant above the 1 per cent level by the Student's *t* test for paired data. A statistically significant, but slight, change toward greater upper-level control took place in the Hierarchically-controlled program.)

Both Tables 1 and 2 show that the clerks in the Autonomy program perceive as predicted a significant shift away from upper-level control when their before-after answers are compared, and that the clerks in the Hierarchically-controlled program see some increase in upper-level control over policy-making, even though it was already perceived as highly controlled from above before the experiment.

These measures of successful experimental manipulation suggest that the conditions in the two programs are sufficiently different to permit tests of the experimental hypotheses.

Hypothesis I

This hypothesis states that an increase in the decision-making role of individuals results in increased satisfactions, while a decrease in opportunity for decision-making is followed by decreased satisfaction. The general hypothesis was tested for a variety of specific areas of satisfaction. The attitudinal areas to be reported include: (*a*) self-actualization and growth, (*b*) satisfaction with supervisors, (*c*) liking for working for the company, (*d*) job satisfaction, (*e*) liking for program. Student's one-tailed *t* test for paired data was used for tests of significance. Results reaching the 5 per cent level or above are considered significant.

TABLE 2

EFFECT OF CHANGE PROGRAMS ON PERCEPTION OF DECISION-MAKING ALLOCATION

Experimental Groups	Index of Perceived Decision-Making Allocation				
	Before Mean	After Mean	Diff	*SE* Diff.	*N*
Program I					
Div. A	5.69	4.39	−1.30**	.24	61
Div. B	6.49	4.08	−2.41**	.26	57
Average	6.08	4.24	−1.84**	.18	118
Program II					
Div. C	6.15	6.87	+ .72**	.22	44
Div. D	6.78	7.13	+ .35	.26	44
Average	6.41	7.00	+ .59**	.17	88

NOTE:—Higher values correspond to perception of predominance of upper levels of organization in decision-making.

** Significant at the 1% level.

TABLE 3

EFFECT OF CHANGE PROGRAMS ON FEELINGS OF SELF-ACTUALIZATION ON JOB

Experimental Groups	Index of Perceived Self-Actualization				
	Mean Before	Mean After	Diff.	*SE* Diff.	*N*
Program I					
Div. A	2.67	2.74	+.07	.09	52
Div. B	2.18	2.39	+.21*	.11	47
Average	2.43	2.57	+.14*	.07	99
Program II					
Div. C	2.43	2.24	−.19	.14	43
Div. D	2.30	2.23	−.07	.10	38
Average	2.37	2.24	−.13*	.07	81

NOTE:—Scale runs from 1, low degree of self-actualization to 5, a high degree.

* Significant at the 5% level, one-tailed *t* test for paired data.

Self-actualization. One of the hypotheses of the study was that greater opportunity for regulating and controlling their own activities within the company structure would increase the degree to which individuals could express their various and diverse needs and could move in the direction of fully exploiting their potentialities. An increase in upper-management control on the other hand was predicted to decrease the opportunities for employee self-actualization and growth.

Five questions were used to measure this area: 1, Is your job a real challenge to what you think you can do? 2, How much chance does your job give you to learn things you're interested in? 3, Are the things you're learning in your job helping to train you for a better job in the company? 4, How much chance do you have to try out your ideas on the job? 5, How much does your job give you a chance to do the things you're best at? These five items, which were answered by checking one position on a five-point scale, were intercorrelated and then combined to form an index.[4] Table 3

[4] The items were intercorrelated by the tetrachoric

shows the means for the four divisions and two groups on the self-actualization and growth index.

While both groups of clerks indicated that their jobs throughout the course of the experiment did not give them a very high degree of self-actualization, the experimental programs produced significant changes. In the Autonomy program, self-actualization increased significantly from before to after, and a corresponding decrease was shown in the Hierarchically-controlled program. At the end of the experimental period, the Autonomy program is significantly higher on this variable than the Hierarchically-controlled program.

Satisfaction with supervision. A variety of indices were developed in order to test the hypothesis that the Autonomy program would improve satisfactions with supervisors and that the Hierarchically-controlled program would reduce such satisfactions. Two general types of attitudes were separately measured: (*a*) satisfaction with relations with supervisors and (*b*) satisfaction with supervisors as a representative. These two types of attitudes were studied before and after the experimental period with respect to three levels of supervision: the first-line supervisor, the assistant manager of the division, and the manager of the division. The following three questions were asked for each of these levels in order to tap the clerks' degree of satisfaction with relations with supervisors:

1. How good is your supervisor (assistant manager, manager) at handling people?

2. Can you count on having good relations with your supervisor (assistant manager, manager) under all circumstances?

3. In general, how well do you like your supervisor (assistant manager, manager) as a person to work with?

These three questions were combined to form indices of satisfaction with relations with supervisors, assistant manager, and manager. (The items were intercorrelated for the satisfaction with relations with supervisor index. Through converting to *z* scores, the average intercorrelation of items is found to be .78. Correcting for length of test, i.e., using three items to form the index rather than one, the reliability index is .91 with an *N* of 360.)

Table 4 shows that in general there was a shift toward greater satisfaction with supervisors in the Autonomy program and toward less satisfaction with supervisors in the Hierarchically-controlled program. The divisions, however, show certain characteristic differences in satisfaction at the outset and shift in the expected direction to different degrees.

Both divisions in the Hierarchically-controlled program show a decrease in satisfaction with the first-line supervisor, although the changes are not statistically significant. The after differences between the Autonomy and the Hierarchically-controlled programs are, however, significant.

Satisfaction with relations with both the assistant manager and the manager increased significantly in the Autonomy program and decreased significantly in the Hierarchically-

TABLE 4

EFFECT OF CHANGE PROGRAMS ON SATISFACTION WITH RELATIONS WITH THREE LEVELS OF SUPERVISION

Experimental Groups	Index of Satisfaction				
	Mean Before	Mean After	Diff.	*SE* Diff.	*N*
Relations with Supervisor					
Program I Div. A	4.18	4.15	—.03	.09	62
Div. B	3.19	3.50	+.31*	.14	54
Average	3.71	3.80	+.09	.08	116
Program II Div. C	3.80	3.67	—.13	.11	46
Div. D	3.43	3.29	—.14	.16	45
Average	3.64	3.48	—.16	.10	91
Relations with Assistant Manager					
Program I Div. A	3.49	3.61	+.12	.12	59
Div. B	3.97	4.11	+.14	.11	53
Average	3.71	3.86	+.15*	.08	112
Program II Div. C	3.80	3.34	—.46**	.12	43
Div. D	3.57	3.22	—.35**	.11	43
Average	3.64	3.28	—.36**	.08	86
Relations with Manager					
Program I Div. A	3.84	4.11	+.27**	.08	62
Div. B	4.04	4.20	+.16*	.09	52
Average	3.93	4.15	+.22**	.06	114
Program II Div. C	3.23	2.59	—.64**	.15	43
Div. D	3.87	3.37	—.50**	.13	40
Average	3.50	3.01	—.49**	.10	83

NOTE:—Degree of Satisfaction with Relations with Supervision: five point scale ranging from 1, low degree of satisfaction to 5, high degree of satisfaction.

* Significant at the 5% level one-tailed *t* test for paired data.

** Significant at the 1% level.

method. When these correlations were converted to *z* scores the average intercorrelation was .62, corrected for length of test, a reliability index of .89 was obtained.

controlled program. Each of the divisions within the groups likewise shifted in the hypothesized directions for the two managerial indices. In the Autonomy program the assistant manager index shifted in the right direction for both divisions, but the changes were not statistically significant when each division was tested separately.

Thus while the employees were generally quite satisfied with their relations with their different supervisors, the experimental programs did have the expected effects of increasing the satisfactions of those in the Autonomy program and decreasing the satisfaction of those in the Hierarchically-controlled program. The effects of the programs appear to be most evident in attitudes toward the managerial level and least marked in attitudes toward the first-line supervisors, probably because the managers occupy the key or pivotal positions in the structure (see Kaye, [3]).

The second type of attitude toward supervisors measured was satisfaction with the supervisors as representatives of the employees. Three questions were asked employees as a measure of this type of satisfaction:

1. How much does your supervisor (assistant manager, manager) go out of her (his) way to help get things for the girls in the section?
2. How effective is she (he) in helping you and the other girls get what you want in your jobs?
3. How much does your supervisor (assistant manager, manager) try to help people in your section get ahead in the company?

These three items were intercorrelated for the attitudes toward the supervisor as a representative index and the average intercorrelation was .83 with a corrected reliability of .94 (N of 340).

The findings for the three levels of supervision on the satisfaction with supervisors as representatives index are shown in Table 5.

The employees' attitudes toward their supervisors as effective representatives of their interests show significant changes in the predicted directions in the two programs. Those in the Autonomy program became more satisfied than they had been previously, while those in the Hierarchically-controlled program became less satisfied. On satisfaction with the first-line supervisor as a representative both Division B in the Autonomy program and Division D in the Hierarchically-controlled program shifted significantly in the hypothesized directions, although the other two divisions did not shift significantly. The two program groups were not matched on degree of satisfaction with manager and assistant manager as a representative at the beginning of the experiment, as there was significantly more satisfaction in the Autonomy program divisions than there was in Program II. However, the changes for both groups of divisions were statistically significant and in the predicted direction. For attitude toward manager all of the division differences are in the predicted direction and all except Division D are statistically significant.

TABLE 5

THE EFFECT OF CHANGE PROGRAMS ON SATISFACTION WITH THREE LEVELS OF SUPERVISION AS REPRESENTATIVES OF EMPLOYEES

Experimental Groups	Index of Satisfaction				
	Mean Before	Mean After	Diff.	*SE* Diff.	*N*
Supervisor as Representative of Employees					
Program I Div. A	3.98	4.06	+.08	.12	59
Div. B	2.91	3.43	+.52**	.14	49
Average	3.48	3.74	+.26**	.09	108
Program II Div. C	3.73	3.67	−.06	.13	45
Div. D	3.52	3.16	−.36*	.18	41
Average	3.39	3.43	−.16	.11	86
Assistant Manager as Representative of Employees					
Program I Div. A	3.32	3.75	+.43**	.14	51
Div. B	3.54	3.76	+.22*	.13	53
Average	3.43	3.75	+.32**	.09	104
Program II Div. C	3.07	2.81	−.26*	.12	41
Div. D	3.23	2.92	−.31*	.13	42
Average	3.15	2.86	−.29**	.10	83
Manager as Representative of Employees					
Program I Div. A	3.82	4.37	+.55**	.11	57
Div. B	3.76	3.96	+.20*	.10	53
Average	3.79	4.17	+.38**	.07	110
Program II Div. C	2.70	2.19	−.51**	.13	41
Div. D	3.14	2.92	−.22	.16	30
Average	2.92	2.52	−.40**	.10	71

NOTE:—Five-point scale ranging from 1, low degree of satisfaction to 5, high degree of satisfaction.

* Significant at the 5% level one-tailed t test for paired data.

** Significant at the 1% level.

Satisfaction with the company. One general question was used to measure company satisfaction: "Taking things as a whole, how do you like working for_____ (the name of the company)?"

The answers for this question presented in

TABLE 6

THE EFFECT OF CHANGE PROGRAMS ON SATISFACTION WITH THE COMPANY

Experimental Groups	Index of Satisfaction with Company				
	Before Mean	After Mean	Diff.	*SE* Diff.	*N*
Program I					
Div. A	4.16	4.32	+.16*	.09	62
Div. B	3.83	4.02	+.19	.13	53
Average	4.01	4.18	+.17*	.08	115
Program II					
Div. C	4.04	3.80	−.24*	.14	46
Div. D	4.26	3.95	−.31**	.12	43
Average	4.15	3.88	−.27**	.09	89

NOTE:—Five point scale, ranging from 1, low degree of satisfaction to 5, high degree of satisfaction.

* Significant at the 5% level one-tailed *t* test for paired data.

** Significant at the 1% level.

TABLE 7

THE EFFECT OF CHANGE PROGRAMS ON JOB SATISFACTION

Experimental Groups	Index of Job Satisfaction				
	Before Mean	After Mean	Diff.	*SE* Diff.	*N*
Program I					
Div. A	3.29	3.29	0	.08	58
Div. B	3.03	3.09	+.06	.09	55
Average	3.16	3.19	+.03	.06	113
Program II					
Div. C	3.14	2.94	−.20*	.10	42
Div. D	3.12	3.07	−.05	.12	46
Average	3.13	3.00	−.13*	.07	88

NOTE:—A five-point scale ranging from 1, low degree of satisfaction to 5, a high degree of satisfaction.

* Significant at the 5% level one-tailed *t* test for paired data.

Table 6 indicate an increase in favorableness toward the company under the Autonomy program and a decrease under the Hierarchically-controlled program.

All of the changes are significant in the predicted direction, except for the before-after difference in Division B which is only at the 10 per cent level of significance.

Job satisfaction. Three questions were used as an index of job satisfaction:

1. Does your job ever get monotonous?
2. How important do you feel your job is compared with other jobs at (the company)?
3. In general, how well do you like the sort of work you're doing in your job?

These three questions showed an average intercorrelation of .47 with a corrected reliability of .73 (*N* of 369). The results on this index are reported in Table 7.

While the trend for the changes in job satisfaction are in the direction predicted, the differences are not sufficiently great to be statistically significant except for Division C. The lack of change in job satisfaction in the Autonomy program may be due to the fact that the job content remained about the same. It is also possible that the increases in complexity and variety of their total work were offset by a rise in their level of aspiration, so that they expected more interesting and varied work.

Satisfaction with the program. In the after measurement additional questions were asked concerning attitudes toward the programs. Most of these questions were open-ended and required the employee to write her response in her own words. Although less than half of the clerks taking the after measurement filled them out, the results on questions relevant to the satisfaction hypothesis deserve brief mention. The clerks in the Autonomy program typically: wanted their program to last indefinitely, did not like the other program, felt that the clerks were one of the groups gaining the most from the program and described both positive and negative changes in interpersonal relations among the girls. The clerks in the Hierarchically-controlled program, on the other hand, most frequently: wanted their program to end immediately, liked the other program and felt that the company gained the most from their program. Not one single person in the Hierarchically-controlled program mentioned an improvement in interpersonal relations as a result of this program. All of the noted changes were for the worse, with increases in friction and tension being most frequently mentioned.

Taking all of these results on the attitudinal questions together, the first hypothesis would appear to be verified. Increasing local decision-making increased satisfaction, while decreasing the role of rank-and-file members of the organization in the decision-making decreased it.

Hypothesis II

This hypothesis predicts a direct relationship between degree of rank-and-file decision-making and productivity. Thus, in order for the hypothesis to be verified, productivity should increase significantly in the Autonomy

program, and should decrease significantly in the Hierarchically-controlled program.

We have previously described the problems of assuming a direct relationship between motivation to produce and productivity in a situation in which volume is not controllable by employees and level of productivity depends upon the number of people doing a fixed amount of work. The Autonomy program was handicapped by both the fact that increasing productivity required reducing the size of their own work group and the fact that the upper management staff and line costs were not included in the measure of costs per volume of work.

The measure of productivity, then, is a measure of clerical costs. These clerical costs are expressed in percentage figures, calculated by dividing the actual clerical costs by a constant standard of cost expected for that volume. Since this way of estimating productivity makes the higher figures indicate lower productivity, we have reversed the signs for purposes of presentation. The results for this measure are shown in Table 8.

The clerical costs have gone down in each division and thus productivity has increased. All these increases in productivity are statistically significant (by *t* tests). In addition, the productivity increase in the Hierarchically-controlled program is significantly greater than that in the Autonomy program. These increases in productivity do not seem to be accounted for by a general rise in productivity throughout the company, since the divisions outside the experimental groups which were most comparable to them showed no significant gain in productivity during this period. The rise in productivity appears to be the result of the experimental treatments. The two divisions initially low in productivity showed the greatest differential change. Division D increased its productivity the most of the four while Division A increased the least.

A second measure of the organizational costs of the two programs is the degree of turnover which could be attributed to on-the-job factors. A method of control and regulation which reduces clerical costs, but which produces the hidden costs of training new employees is of greater cost to the organization than would at first appear evident. In this company turnover, however, is not high and much of the turnover

TABLE 8

COMPARISON OF THE FOUR DIVISIONS ON CLERICAL PRODUCTIVITY FOR YEAR CONTROL PERIOD AND YEAR EXPERIMENTAL PERIOD

Experimental Groups	Index of Productivity				
	Mean Control Period	Mean Experimental Period	Diff. %	*SE* Diff.	*N*
Program I					
Div. A	46.3%	55.2%	+ 8.9**	1.3%	12
Div. B	51.0	62.0	+11.0**	1.3	12
Average	48.6	58.6	+10.0**	1.2	24
Program II					
Div. C	50.2	63.2	+13.0**	1.2	12
Div. D	46.8	62.0	+15.2**	1.1	12
Average	48.5	62.6	+14.1**	.9	24

NOTE:—Higher values correspond to greater productivity.
** Significant at the 1% level.

that does occur is due to personal reasons (marriage, pregnancy, etc.) rather than on-the-job reasons. Out of the 54 employees who left the company from the four divisions during the time of the experiment, only nine resigned for other jobs or because of dissatisfaction. Out of these nine, however, all but one were in the Hierarchically-controlled program. In the exit interviews conducted by the company personnel department 23 of the girls leaving made unfavorable comments about pressure, work standards, etc. Nineteen of these girls were from the Hierarchically-controlled program.

These results indicate that the productivity hypothesis is clearly not verified in terms of direct clerical costs, since the Hierarchically-controlled program decreased these costs more than the Autonomy program, contrary to the prediction. The indirect costs for the Hierarchically-controlled program are probably somewhat greater. But even when this is considered the evidence does not support the hypothesis.

DISCUSSION

The results on productivity might suggest a "Hawthorne effect" if it were not for the satisfaction findings. The increase in satisfaction under the Autonomy program and the decrease under the Hierarchically-controlled program make an explanation of productivity changes in terms of a common attention effect unlikely.[5]

[5] It is unlikely that even in the Hawthorne experi-

The Hierarchically-controlled program reduced staff costs by ordering reductions in the number of employees assigned to the tasks. Increases in productivity for Divisions C and D were brought about as simply as that. This temporary increase in one measure of productivity is not surprising and is traditional history in industry. In the Autonomy program, decrease in costs was more complex but can be simply stated as follows. The Autonomy program increased the motivation of the employees to produce and thus they did not feel the need for replacing the staff members who left the section. In addition, they were willing to make an effort to try to outplace some of their members in other jobs which they might like. The reductions in staff in the two programs came about in different ways. Those occurring by order in the Hierarchically-controlled program surpassed in number those occurring by group decision in the Autonomy program, but it is not clear how long the superiority of the Hierarchically-controlled program would have lasted.

The results of the experiment need to be placed in a larger theoretical framework in order to contribute to the understanding of the functioning of large-scale organizations. We shall first consider briefly the role and function of the social control processes, as it is these processes which were changed by the experimental manipulations.

The high degree of rationality which is characteristic of the institutional behavior of man is achieved through a complex system for controlling and regulating human behavior. Hierarchy is a requirement because human beings must be fitted to a rational model. There are essentially two functions which the usual hierarchy serves: a *binding-in* function and a *binding-between* function. By *binding-in* we mean insuring that there will be individuals present to fill the necessary roles. The role behavior required by the organization must be a path to individual goals. Money is the most important means used for binding-in, but all ways to motivate a person to enter and remain in the system are means of binding-in. By *binding-between* we mean the insurance of the rationality of action, that is, the setting up and continuation of institutional processes which will accomplish the ends for which the organization is designed. The role behavior of individuals must be integrated into a pattern to produce interrelated action directed toward the goals of the organization. The development of assignments, work charts, job specifications, etc., are but a few examples of the many means used by organizations for binding-between.

Any means for controlling and regulating human behavior in a large organizational setting, then, needs to serve these two functions. The experiment shows that the allocation of decision-making processes to the upper hierarchy results in a greater emphasis on the binding-between function, while the function of binding-in is handled by an external reward system. Such a direct stress on the binding-between function was shown in the Hierarchically-controlled program and resulted in the increase in productivity (an indication of binding-between) and a decrease in employee satisfaction (an indication of degree of binding-in) and some increase in turnover (another indication of binding-in).

The greater allocation of the decision-making processes to the rank-and-file employees in the Autonomy program resulted in an emphasis on both the binding-between and the binding-in functions. Thus there was both an increase in productivity and an increase in satisfaction. While the program is addressed primarily to the binding-in function, in such a context the binding-between function is also served.

The problems of the Hierarchically-controlled system are maintaining the employee effectively "bound-in" to the organization and continuing favorable relations between the supervisory personnel who have involvement in the organization and the rank and file who must do the work. Indications of these problems are dissatisfaction, distortions in communications up the hierarchy, the tendency to "goof off" and cut corners in the work, and the greater turnover.

The Autonomy program is an integrated means of handling both the binding-between and the binding-in functions, but it requires in the long run that the organization be willing to grant employee decision-making in the key areas of binding-in such as pay and promo-

---

ment the results were due to attention. There were a number of changes in addition to an increase in attention, including relaxation of rules, better supervisors, no change in piece rates despite raises in productivity—to name a few.

tions. The granting of "safe" areas of decision-making and the withholding of "hot" ones is not likely to work for long. It is necessary for the rank and file to be sufficiently bound in to the organization for them to want to make decisions which are rational for the system. But the rationality of their decisions will also depend upon the orientation of the key supervisors whose values they will interiorize. (Thus the clerks in Division B were more organizationally oriented than those in Division A—see Kaye [3].)

SUMMARY

A field experiment in an industrial setting was conducted in order to test hypotheses concerning the relationship between the means by which organizational decisions are made and (*a*) individual satisfaction, and (*b*) productivity.

Using four parallel divisions of the clerical operations of an organization, two programs of changes were introduced. One program, the Autonomy program involving two of the divisions, was designed to increase the role of the rank-and-file employees in the decision-making processes of the organization. The other two divisions received a program designed to increase the role of upper management in the decision-making processes (the Hierarchically-controlled program). The phases of the experiment included: (*a*) before measurement, (*b*) training programs for supervisory personnel lasting approximately 6 months, (*c*) an operations period of a year for the two experimental programs, and (*d*) after measurement. In addition, certain measurements were taken during the training and operational phases of the experiment. Findings are reported on the question of the experimental "take" and on the general hypotheses on individual satsfactions and productivity. Briefly, it was found that:

1. The experimental programs produced changes in decision-making allocations in the direction required for the testing of the hypotheses.

2. The individual satisfactions of the members of the work groups increased significantly in the Autonomous program and decreased significantly in the Hierarchically-controlled program.

3. Using one measure of productivity, both decision-making systems increased productivity, with the Hierarchically-controlled program resulting in a greater increase.

The relationship of the findings to the so-called "Hawthorne effect" is examined and the experimental programs and their results are considered in the light of a theoretical description of the role of the control and regulation processes of large organizations.

REFERENCES

1. COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512–532.
2. KATZ, D., MACCOBY, N., & MORSE, NANCY. *Productivity, supervision and morale in an office situation.* Ann Arbor: Survey Res. Center, Univer. of Michigan, 1950.
3. KAYE, CAROL. *The effect on organizational goal achievement of a change in the structure of roles.* Ann Arbor: Survey Res. Center, 1954 (mimeographed).
4. LEWIN, K. Group decisions and social change. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.). *Readings in social psychology* (2nd Ed.). New York: Holt, 1952, 459–473.
5. LIPPITT, R., & WHITE, R. K. An experimental study of leadership and group life. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.). *Readings in social psychology* (2nd Ed.) New York: Holt, 1952, 340–354.
6. REIMER, E. *Creating experimental social change in an organization.* Ann Arbor: Survey Res. Center, 1954 (mimeographed).
7. TANNENBAUM, A. *The relationship between personality variables and adjustment to contrasting types of social structure.* Ann Arbor: Survey Res. Center 1954 (mimeographed).
8. WORTHY, J. C. Factors influencing employee morale. *Harvard Bus. Rev.*, 1950, **28**, 61–73.

Received September 19, 1955. Prior publication.

CORRELATES OF TWO TYPES OF XENOPHILIC ORIENTATION[1]

HOWARD V. PERLMUTTER

Center for International Studies, Massachusetts Institute of Technology

IN RECENT years there has been an interest in qualifying the first findings of research on the authoritarian personality (1). Early studies usually reported a high correlation between authoritarianism, ethnocentrism, and xenophobia. It does not follow, however, that low authoritarians necessarily exhibit *xenophilia*, that is, a preference for foreign people and their ways of life. In fact, to the extent that both xenophilia and xenophobia involve a tendency to draw sharp distinctions between ingroups and outgroups, these orientations are equally compatible with authoritarianism. In a recent study (3) it was reported that individuals scoring high on a scale devised to measure xenophilia also scored higher on the F scale than those rated low on xenophilia. Subsequent studies have corroborated this general relationship.

THEORETICAL CONSIDERATIONS

These studies raise problems concerning the psychological significance of the xenophilic syndrome.

1. It is logically possible for an individual to agree with assertions favoring a foreign way of life over his own without experiencing or expressing strong hostility toward his own country. Do individuals scoring high on xenophilia express hostility toward their compatriots more openly than those scoring low on xenophilia?

2. Subjects in the previous study had not been given opportunity to indicate when they saw no difference between nationals of one country and those of another. In a situation in which unwillingness to stereotype is an acceptable answer are high xenophiles more willing to stereotype than low xenophiles?

3. It is possible that people indicate attraction toward foreign institutions without desiring to visit or live in foreign countries. Are high xenophiles more disposed than low xenophiles to travel or live abroad, and if so, how strong are these motives in high xenophiles as compared to low xenophiles?

4. Although high xenophiles tend to agree with authoritarian items, do the high authoritarians of the high xenophile group differ from the high authoritarians of the low xenophile group? In what ways? The same questions may be asked about low authoritarians in the two groups.

Clinical studies of individuals scoring high and low on xenophilia aided in formulating a picture of four types of xenophilic orientation: the low authoritarian-high xenophiles and high authoritarian-high xenophiles were distinguished from each other as well as from the low authoritarian-low xenophiles and high authoritarian-low xenophiles in this way:[2]

1. *The low authoritarian-high xenophile* is considered to be the most strongly attracted to foreign people and ways of life. His xenophilic orientation is based as much on a rejection of ingroup authorities, as on an identification with the foreign outgroup. He is most likely then to express hostility toward his own country, most likely to be concerned with travel to another country, and most likely to make generalizations about the positive attributes of foreigners and the negative attributes of domestic people.

2. *The high authoritarian-high xenophile* is thought to resemble in many ways the authoritarian personality previously described (1). But his ambivalent solution to ingroup-outgroup problems involves unconsciously rejecting the ingroup and favoring the outgroup while maintaining his basic tendencies not to

[1] This is part of a program on Personality Determinants in International Communication being conducted in the International Communications Program at the Center for International Studies, Massachusetts Institute of Technology. The author wishes to express his sincere thanks to Dr. M. Brewster Smith of the Social Science Research Council, Dr. Henry Weinberg, Department of Psychology, Boston University, Dr. Ithiel de Sola Pool, Mr. Frank Bonilla and Mrs. Doris Held at the Center for International Studies for their criticisms and suggestions.

[2] Of particular importance were two unpublished papers: Perlmutter, H.V. & Weinberg, H., Xenophilia and Authoritarianism: Two Case Studies, and Held, D., Values and Attitudes: An Exploratory Study Into Their Differing Functions for Personality, both available at the Center for International Studies, Massachusetts Institute of Technology, Cambridge, Massachusetts.

express disfavor of authorities, either foreign or domestic. He is more reluctant to express hostility toward his own country and sever connections with his own country than other high xenophiles.

Since holding xenophilic attitudes implies a positive orientation toward a foreign way of life, high xenophiles as a group were presumed to differ from the low xenophiles in a greater readiness to stereotype, a greater willingness to travel and, in general, a greater willingness to express hostility toward their own country. The two kinds of low xenophiles were expected to differ primarily in their willingness to stereotype.

3. *The low authoritarian-low xenophile* is assumed to be more likely to be an "equalitarian" personality. His disagreement with xenophilic items does not imply that the domestic way of life is favored over the foreign. It is expected that he is unwilling to draw comparisons between typical Americans and other foreigners because he sees people as essentially equal or too varied to compare.

4. *For the high authoritarian-low xenophile* it is expected that disagreement with the items means agreement with the converse. These individuals are disposed to believe that the foreign way of life is less desirable than the domestic. In line with research on the authoritarian-ethnocentric syndrome, it seems likely that this group would also be more likely to score higher on willingness to stereotype.

The relationships postulated in this discussion are summarized in the form of specific hypotheses, which are formally stated below with the supporting or negative evidence for each.

PROCEDURE

A total of 422 subjects including 178 males and 244 females between the ages of 19 and 23 at the College of Liberal Arts, College of Business Administration, and College of Practical Arts and Letters at Boston University, were administered a questionnaire which included (*a*) the Xenophile Scale, (*b*) the Authoritarian Scale (revised), (*c*) a scale measuring hostile feelings toward "typical" Americans and (*d*) a questionnaire on desire to travel and live abroad.

Instruments

The *Xenophile Scale* consists of ten items involving comparisons between American, European, and to a lesser extent, Oriental ways of life to the detriment of the American. Subjects' responses to these items may range from +3 (high agreement) to −3 (high disagreement). The test is described completely in a previous article (3). All scores were then transferred to a seven-point Likert-type scale, in which 1 indicated complete disagreement and 7 indicated complete agreement with an item.

The *Authoritarian Scale* consists of 17 items, most of them taken from the F Scale of Adorno *et al.* (1), with some additional items emphasizing admiration for directive leadership.[5] Again subjects' responses, ranging from +3 to −3, were transferred to a seven-point Likert-type scale.

The *Scale of Hostile Feelings toward "Typical" Americans*, previously standardized on a similar population, lists fifteen positive and fifteen negative traits.[6] Subjects were asked to complete the phrase: "In comparison with the typical European, Americans are (more/less)...." Subjects who said Americans possessed more of a negative trait than Europeans were given a score for negative feelings about "typical" Americans; subjects who indicated that "typical" Americans possessed less of a positive trait were also given a score for negative feelings. Both scores are considered indices of hostility toward "typical" Americans.

Also in this test subjects could mark zero if they felt there were no differences between "typical" Americans and "typical" Europeans, thus indicating their unwillingness to stereotype. A high zero score indicates a relatively higher unwillingness to stereotype than a low zero score.

The Scale of Interest in Travel and Residence Abroad, a five-point rating scale measuring various degrees of interest in living abroad, was devised. This scale categorized responses to the following questions:

1. How likely is it that you will actually go abroad?_____

2. I think of living in [a foreign country of first choice] *one* year with all expenses paid._____

3. Suppose you had an opportunity to accept a very

[5] The following F-scale items of forms 45 and 40 were used: 1, 2, 6, 8, 18, 19, 21, 22, 23, 26, 27, 41; and items of form 60 used were: 41 and 44 (1, pp. 248–250, 255–257). In addition the following three items, suggested by J. W. Thibaut and H. W. Riecken were used: 12. In the case of disagreement within a group the judgment of the leader should be final; 20. One quality of a good leader is his ability to stick to his plans even though some of his followers object and try to change his mind; and 27. If the leader of a group pays attention to the individual problems and desires of each member, he is not likely to get much accomplished by the group.

[6] The fifteen positive traits were: imaginative, understanding, cultured, deep, frank, cooperative, democratic, adaptable, poised, broad interests, easy going, mature, self confident, friendly, well mannered.

The fifteen negative traits were: dishonest, frivolous, worrying, hostile, autocratic, neurotic, pessimistic, selfish, cold, naive, passive, unstable, prejudiced, untrustworthy, materialistic.

A sample of 100 students rated a longer list of traits with respect to the relative desirability or undesirability of each trait. These 15 positive traits were seen as significantly "desirable" while the 15 negative traits were seen as significantly "undesirable."

TABLE 1

CHARACTERISTICS OF GROUPS CATEGORIZED BY XENOPHILIA AND AUTHORITARIANISM

Group	*N*	Scale Scores			
		Xenophilia		Authoritarianism	
		Mean	*SD*	Mean	*SD*
Combined high xenophiles	144	49.3	4.86	66.4	14.21
High xenophiles, high authoritarians	53	49.3	3.97	79.7	6.01
High xenophiles, low authoritarians	42	50.0	6.22	48.3	9.06
Combined low xenophiles	142	26.7	4.40	59.6	15.33
Low xenophiles, high authoritarians	31	26.9	4.84	79.8	7.71
Low xenophiles, low authoritarians	61	27.2	5.30	48.7	9.70

desirable job requiring that you live over ten years abroad and that you become a citizen of the foreign country in which you are working—how likely would you be to accept the job?_____

4. If going abroad meant living under very *limited* financial resources and meant a great number of personal sacrifices—how likely is it that you would go?_____

Classification of Subjects

All subjects who, on the Xenophile Scale, scored in the upper third of the distribution were termed high xenophiles. These numbered 144 of whom 58 were males and 86 were females. All those who fell in the lower third of the distribution, were termed low xenophiles. This group numbered 142, including 60 males and 82 females.

F-scale scores were obtained for both the high and low xenophiles. Within each of the two groups, the top third in the distribution of authoritarianism scores was designated high authoritarian and the lower third designated low authoritarian. Those with intermediate scores were discarded. This yielded four groups for analysis: (*a*) Those scoring high on xenophilia and authoritarianism, (*b*) those high on xenophilia but low on authoritarianism, (*c*) those low on xenophilia and high on authoritarianism and (*d*) those who scored low on both dimensions. Means and standard deviations on the authoritarianism and xenophilia scales are given for these four groups in Table 1.

RESULTS

Hypothesis I

High scorers on the Xenophile Scale display a greater tendency to express hostility toward "typical" Americans than low scorers on the Xenophile Scale.

TABLE 2

HOSTILITY TOWARD TYPICAL AMERICANS IN RELATION TO XENOPHILIA AND AUTHORITARIANISM

Standing on Xenophilia and Authoritarianism	*N*	Hostility Toward Typical Americans					
		Americans More: Negative Traits		Americans Less: Positive Traits		Zero Score	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
Combined high xenophiles	144	5.8	2.69	5.2	2.59	8.2	4.70
High xenophiles, high authoritarians	53	5.8	2.65	4.5	2.35	8.6	4.62
High xenophiles, low authoritarians	42	6.1	3.00	6.2	2.78	8.1	4.79
Combined low xenophiles	142	4.1	1.97	3.1	2.16	10.9	5.20
Low xenophiles, high authoritarians	31	4.5	2.06	3.1	2.48	9.1	2.83
Low xenophiles, low authoritarians	61	3.8	1.95	3.1	2.16	12.6	6.01

Table 2 indicates that Hypothesis I is supported. The high xenophiles score significantly higher than the low xenophiles on both indices of hostile feelings. They report that typical Americans possess more negative traits and fewer positive traits.

Hypothesis Ia

Among the high xenophiles, low scorers on the F scale display a greater tendency to express hostility toward "typical" Americans than high scorers on the F scale.

This hypothesis also receives some support. Low authoritarians among the high xenophiles attribute significantly fewer positive traits to typical Americans than do high authoritarians who are high xenophiles. These low authoritarians also have a tendency to attribute more negative traits to typical Americans than the low authoritarians, although the difference is not significant.[5]

[5] To indicate differences between means of responses along the various dimensions the *t* test was used. A difference between two means was regarded as significant when *p* was less than .01 level. In a few instances where the trends in the data are especially strong and consistent with our hypothesis, the *t* ratio is reported. In all cases a one-tailed *t* test was used.

Hypothesis Ib

Among the low xenophiles, low authoritarians have a greater tendency to express hostility toward typical Americans than high authoritarians.

The results in Table 2 offer no confirmation of this hypothesis. Actually there is a slight tendency for high authoritarians to attribute more negative traits to typical Americans than the low authoritarians. This difference is not significant.

Hypothesis II

High scorers on the Xenophile Scale display a greater tendency to "stereotype" Americans and "typical Europeans" than low scorers on the Xenophile Scale.

Table 2 indicates a significant difference in the predicted direction. The low xenophile group combined has a higher zero score than the high xenophile group, thus indicating a greater unwillingness to draw stereotypic comparisons between Americans and Europeans.

Hypothesis IIa

Among the high xenophiles, the low authoritarians display the tendency to stereotype to a greater degree than the high authoritarians.

A significant difference was not found, but the difference is in the predicted direction. The low authoritarians among the high xenophiles make stereotypic judgments somewhat more readily than the high authoritarians of the high xenophile group.

Hypothesis IIb

Among the low xenophiles, the high authoritarians display a greater tendency to stereotype than the low authoritarians.

This hypothesis is unequivocally supported by the data. The difference between the two groups is clearly significant.

Hypothesis III

High scorers on the Xenophile Scale are more interested in travel and residence abroad than low scorers on the Xenophile Scale.

The results relevant to Hypothesis III are presented in Table 3. Hypothesis III is supported in all comparisons:

1. High xenophiles feel it is somewhat more likely they will go abroad than low xenophiles ($t = 1.5$).
2. High xenophiles are significantly more likely to think of living abroad than low xenophiles.
3. High xenophiles are somewhat more willing to accept a job overseas that means giving up citizenship than low xenophiles ($t = 1.9$).
4. High xenophiles would be more willing to live abroad under difficult economic circumstances than low xenophiles. This difference is significant.

TABLE 3

DESIRE TO LIVE ABROAD IN RELATION TO XENOPHILIA AND AUTHORITARIANISM

(A rating of one indicates high desire; a rating of five indicates low desire)

Standing on Xenophilia and Authoritarianism	*N*	Desire to Live Abroad							
		Likely to Go Abroad		Think of Going Abroad		Live Abroad Ten Years; Give Up Citizenship		Live in Poverty Abroad	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Combined high xenophiles	95	2.8	0.91	3.0	1.00	3.4	1.12	3.0	0.97
High xenophiles, high authoritarians	38	2.9	0.79	2.9	1.07	3.6	1.09	3.2	0.98
High xenophiles, low authoritarians	27	2.5	0.97	2.7	1.00	2.9	1.16	2.6	0.91
Combined low xenophiles	100	3.0	0.94	3.4	1.12	3.9	0.93	3.5	1.05
Low xenophiles, high authoritarians	23	3.0	0.82	3.4	1.16	3.8	1.00	3.9	0.92
Low xenophiles, low authoritarians	42	2.9	0.91	3.4	1.04	3.9	0.86	3.4	1.25

NOTE—Not all subjects filled out identical forms of the Desire to Travel Questionnaire. Only those who completed identical scales were used in this table.

Hypothesis IIIa

Among high xenophiles, low authoritarians are more concerned with living abroad than the high authoritarians.

On all four indices of the desire to travel, the trend of the data seems to support this hypothesis. The significant differences occur with respect to the two extreme conditions, given in the two right-hand columns in Table 3. The low authoritarian-high xenophiles report more often that they are willing to give up citizenship in their own country in order to gain economic advantages. They are also more ready to accept economic deprivation in exchange for an opportunity to live abroad.

Hypothesis IIIb

Among the low xenophiles, the high authoritarians are less often disposed to live and travel abroad than low authoritarians.

From Table 3, it is clear that little data can be found to support Hypothesis IIIb. There is a tendency, however, for the high authoritarians to be less disposed to live abroad under disadvantageous economic situations ($t = 1.8$).

Hypothesis IIIc

High authoritarians among the high xenophiles are more disposed to live abroad, etc., than high authoritarians among the low xenophiles.

While in all of the four comparisons differences are in the predicted direction, only willingness to live in poverty abroad distinguishes the high authoritarians in one xenophilic category from the other at a significant level. The high authoritarians among the high xenophiles seem to be more disposed to live abroad even under economically difficult circumstances than the high authoritarians among the low xenophiles.

Hypothesis IIId

Low authoritarians among the high xenophiles display greater tendencies to live abroad than low authoritarians among the low xenophiles.

Differences are in the predicted direction in every case. For the two extreme conditions (right-hand columns in Table 3) the differences are significant and most striking.

DISCUSSION

In general the results indicate that the typology based on responses to a scale designed to measure attraction to things foreign and the California F scale warrants further research.

High Xenophiles and Low Xenophiles

The differences found between high and low scorers on the Xenophile Scale with respect to the dependent variables of hostility toward the typical American, willingness to stereotype, and desire to travel suggest that the scale effectively locates individuals along a meaningful dimension of attitude. In previous researches (2, 3) the question had been raised whether high xenophiles and high authoritarians share a general disposition to agree with dogmatic statements irrespective of their content. To score high on the Xenophile Scale an individual must agree with most of the ten items given. But to score high on the scale of expression of hostility simple agreement with dogmatic statements is not involved; the subject had to say that (*a*) typical Americans possess more negative traits or (*b*) typical Americans possess fewer positive traits. The present data show that high xenophiles on the average express positive and negative statements consistent with their orientation toward foreign people.

The High Authoritarian-High Xenophile Orientation

In spite of his disposition to defer to authority, the high authoritarian-high xenophile seems particularly ambivalent about domestic authority figures. While overtly advocating authority and hierarchy, he seeks to escape it. Thus, in fantasy he can deal with his fear of domestic authority figures by moving toward less threatening foreign ones and find satisfaction in Europe's "tradition orientation" and social hierarchy. Yet because he is ambivalent, he is less willing than the low authoritarian-high xenophile to give up his citizenship in the home country.

But it has not been demonstrated that the high authoritarian-high xenophile is necessarily attracted to all "foreign" groups. It may be that both the high authoritarian-high xenophile and the high authoritarian-low xenophile share a dislike and fear for "weak" domestic minority groups, but the high authoritarian-high xenophile differs from the high authoritarian-low xenophile in that he admires the foreign way of life in some foreign countries. In

short, he can be an "Anglophile" or a "Francophile" and dislike Jews, Negroes, and Mexicans. This interpretation remains to be tested.

The Low Authoritarian-High Xenophile Orientation

The data suggest that the low authoritarian-high xenophiles appear to be far more involved in the xenophilic orientation. Their interest in living abroad under unfavorable conditions and the willingness of some to give up citizenship in their own country suggest that their low F scale scores could be interpreted in two different ways. It is possible that these low authoritarians are nonauthoritarian xenophiles. That is, their reaction to authority figures is flexible, relatively objective, and without anxiety or fear. Or it is possible that they are antiauthoritarian, that is, they have strong dispositions to disagree with, rebel against, and escape from authorities. The latter interpretation seems more plausible. Since the subjects studied were college students between the ages of 19 and 23, presumably they are still involved in the resolution of adolescent conflicts concerning parental authority. A fear of being controlled by the domestic authorities and a need to assert freedom from parental influence makes the foreign situation inviting and results in their positive orientation toward things foreign. But further research is necessary to distinguish the nonauthoritarian and antiauthoritarian syndromes.[6]

[6] Research is currently under way at the Center for International Studies to develop scales and criteria for antiauthoritarianism. In addition, there is concurrently an attempt to study the differences between the xenophilic orientation, wherein identification is presumed to be with the foreign, and the "cosmopolitan" orientation, wherein identification remains primarily with the domestic but attraction is expressed to aspects of the foreign.

SUMMARY

A sample of 422 college students at Boston University responded to scales measuring tendencies toward xenophilia, authoritarianism, hostility toward typical Americans, willingness to stereotype, and interest in travel abroad.

The results, on the whole, support the hypotheses which the study sought to test. High scorers on the Xenophile Scale can be distinguished from low scorers on the above dimensions. High xenophiles seem to be more willing to stereotype, express hostility toward typical Americans, and exhibit a stronger desire to travel than low xenophiles.

Predictions were made concerning two varieties of high xenophiles, high scorers and low scorers on the California F scale. It was found that the low authoritarians were more attracted to things foreign than the highs along several dimensions.

Furthermore among low scorers on the Xenophile Scale, clear distinctions were found among the high and low F scale scorers. The highs were more willing to stereotype than the lows.

In addition, significant differences were found between the high authoritarians of the high xenophile and low xenophile groups, and the low authoritarians of the high and low xenophile groups.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. COHN, T. S. The relation of the F-scale to a response to answer positively. *Amer. Psychologist*, 1953, **8**, 335. (Abstract)
3. PERLMUTTER, H. V. Some characteristics of the xenophilic personality. *J. Psychol.*, 1954, **38** 291–300.

Received October 25, 1955. Prior publication.

CRITIQUE AND NOTES

THE EFFECTS OF MANIFEST ANXIETY AND IQ ON DISCRIMINATION

JEAN S. KERRICK[1]

Department of Journalism, University of California

EXPERIMENTAL evidence indicates that both intelligence and manifest anxiety are significantly related to precision of judgment or discrimination. However, seemingly conflicting evidence has been reported concerning the nature of the relationship of manifest anxiety to conditioned discrimination.

Two experiments, both differentiating subjects (*Ss*) on the Taylor Scale of Manifest Anxiety (7), have been concerned with simple and then with discriminatory conditioning. Hilgard, Jones, and Kaplan (2), report correlations approaching significance between anxiety scores and lack of discrimination (the ratio of negative to positive CR's and the frequency of negative R's). On the other hand, Spence and Beecroft (6), considering reaction time, found excitatory strength of the positive CS was significantly greater for highly anxious *Ss*. The experiments both used college students as *Ss*.

Hilgard, Jones, and Kaplan suggest that highly anxious *Ss* show a lack of discrimination. We would expect, then, that from a graduated series of responses, they would be more likely to utilize the extreme positions rather than to make finer distinctions among stimuli.

Osgood (4) has established that, in discrimination along a series, a shorter reaction time is associated with extreme judgments. Precision of judgment, rather than an "all-or-none" response, should therefore be associated with longer reaction-time.

Spence and Beecroft, measuring reaction time, indicate that highly anxious *Ss* are more discriminative. The shorter latency period indicates that anxious *Ss* are more readily able to make "either-or" distinctions. This, according to Osgood's data, would suggest that they would make less fine distinctions when presented with a gradient. The conflict between the two sets of data, then, may be explained by the differences in the measure of discrimination and by Osgood's finding that the factor of intensity, rather than of direction, is crucial to reaction-time.

If we accept this explanation we should expect that, given stimuli to judge on a 7-point scale, highly anxious *Ss* would use the extreme positions and the neutral position more often than the *Ss* who are minimally anxious.

That IQ is related to discrimination along a gradient appears obvious in terms of the definition of the two terms. Also, Osgood reports that a preliminary analysis of his data indicated that differences in the use of scale positions are related to education and intelligence, the more critical thinkers making a more discriminatory use of the entire scale.

The evidence indicates two hypotheses: (*a*) high-anxiety *Ss* will make less fine discriminations than low-anxiety *Ss*, and (*b*) high IQ *Ss* will make finer discriminations than low IQ *Ss*.

Bindra, Paterson, and Strzelecki (1) attempt to clarify the explanation of the effects of anxiety in a conditioned salivation experiment. They demonstrate no significant difference between the performance of anxious and nonanxious *Ss*. They interpret their data to be in accord with Hilgard, *et al.*, who suggest that anxious *Ss* have a higher drive level only with regard to specific, defensive responses. If this interpretation be true, we would expect to find no difference in the performance of anxious and nonanxious *Ss* on a discrimination task of the present kind.

METHOD

Over 400 high school students were initially given the Advanced Examination of the Otis IQ test and the Taylor Scale of Manifest Anxiety. Classes from freshman to senior, half male and half female, were used. Distributions of scores on both the IQ and the anxiety tests were approximately normal.

Twenty *Ss* were selected so that five of them fell into each of the following groups on the basis of quartiles of IQ and anxiety: (*a*) high IQ, high anxiety, (*b*) high IQ, low anxiety, (*c*) low IQ, high anxiety, and (*d*) low IQ, low anxiety.[2]

The *Ss* were given a form of Osgood's semantic differential, a measure of "meaning" (5). The differential, a combination of associational and scaling procedures,

[1] The data reported here were collected as part of a larger study submitted as a dissertation for the degree Doctor of Philosophy in mass communications at the University of Illinois. The supervision and advice of Dr. Charles E. Osgood is gratefully acknowledged. Funds for the study were made available by the Institute of Communications Research at the University of Illinois.

[2] The upper and lower quartiles of the two distributions were used. High IQ's ranged from 106–127, low IQ's from 75–92. High-anxiety scores ranged from 25–44, low ranged from 2–13.

TABLE 1
USE OF SCALES BY HIGH AND LOW ANXIETY SUBJECTS

	Scale Positions			
	+3, −3	+2, −2	+1, −1	0
High Anxiety	438	166	123	273
Low Anxiety	427	166	119	288
Total	865	332	242	561

$\chi^2 = .610$, $df = 3$, $p > .10$

TABLE 2
USE OF SCALES BY HIGH AND LOW IQ SUBJECTS

	Scale Positions			
	+3, −3	+2, −2	+1, −1	0
High IQ	390	178	133	299
Low IQ	475	154	109	262
Total	865	332	242	561

$\chi^2 = 14.908$, $df = 3$, $p < .01$

consists of the judgment of a concept against a series of descriptive scales. Each of these scales is defined by a pair of polar adjectives and consists of seven intervals.

The *S*s were required to make judgments of each of 10 attitudinal concepts on each of 10 scales.[3] Their responses on each of the scales and on each of the concepts were tabulated to determine the frequency with which each *S* used each step in the scale. To eliminate the effect of direction of attitude, negative and positive steps of equivalent value on the 7-step scale were combined. Thus the degree of discrimination, rather than the positive or negative values of the concepts was considered.

[3] Concepts used were: Television, The American Council of Churches, Communists, The American Medical Association, Eleanor Roosevelt, The Public Health Service, Mental Hygiene Clinics, Comic Books, The Ku Klux Klan, and the Liquor Retailer's Association. Scales used were: fair–unfair, small–large, dirty–clean, active–passive, good–bad, fast–slow, pleasant–unpleasant, tense–relaxed, weak–strong, valuable–worthless.

RESULTS

Manifest anxiety and discrimination. When the frequency of use of each scale position by the high-anxiety group (upper quartile) was compared with the frequency of use by the low-anxiety group (lower quartile), a chi-square analysis revealed no significant difference. These data are presented in Table 1.

IQ and discrimination. When the upper and lower IQ quartiles were compared, the chi-square value was significant ($p < .01$), indicating a difference in the use of the scales. From the data in Table 2, it is apparent that the low IQ *S*s used the extreme scale positions to a significantly greater extent than the high IQ *S*s did. Judgments by the low IQ *S*s tended to be more polarized.

Manifest anxiety and IQ. An indication of the interaction between IQ and anxiety on discrimination is given by a consideration of both variables for each scale position. These data are presented in Table 3. All of the four comparisons are significant. A schematic comparison of the magnitude of use of each position for anxiety groups under conditions of low and high IQ (Table 4) indicates that the hypothesis of an interaction between the two variables is tenable.

TABLE 3
USE OF EACH SCALE BY IQ AND ANXIETY GROUPS

−3, +3

	High IQ	Low IQ	Total
High Anx.	214	224	438
Low Anx.	176	251	427
Total	390	475	

$\chi^2 = 5.098$, $df = 1$, $p < .05$

−2, +2

	High IQ	Low IQ	Total
High Anx.	78	88	166
Low Anx.	100	66	166
Total	178	154	

$\chi^2 = 5.862$, $df = 1$, $p < .05$

−1, +1

	High IQ	Low IQ	Total
High Anx.	45	78	123
Low Anx.	88	31	119
Total	133	109	

$\chi^2 = 34.114$, $df = 1$, $p < .001$

0

	High IQ	Low IQ	Total
High Anx.	163	110	273
Low Anx.	136	152	288
Total	299	262	

$\chi^2 = 8.779$, $df = 1$, $p < .01$

DISCUSSION

Intelligence, as measured by the Otis IQ test, does significantly relate to discrimination. The *S*s

TABLE 4

FREQUENCY OF JUDGMENTS FOR HIGH AND LOW IQ GROUPS IN TERMS OF ANXIETY

Scale Position	High IQ	Low IQ
+3, −3	High anx. > Low anx.	High anx. < Low anx.
+2, −2	High anx. < Low anx.	High anx. > Low anx.
+1, −1	High anx. < Low anx.	High anx. > Low anx.
0	High anx. > Low anx.	High anx. < Low anx.

in the lowest IQ quartile use the extreme scale values (+3 and −3) much more often than *Ss* in the highest IQ quartile. High IQ *Ss* made more use of the whole scale.

High- and low-anxiety *Ss*, on the other hand, make equally fine discriminations along a graduated series. This would seem to contradict the results of both the Hilgard, Jones, and Kaplan experiment, and of the Spence and Beecroft experiment.

However, IQ and anxiety do appear to interact with regard to discrimination. This is consistent with an earlier study (3) which indicated that IQ and anxiety interact to affect scores on a multiple-choice information test. In general, the effect of increased anxiety on high IQ *Ss* is to make them less "discriminating." They use the ±3 positions and the neutral position more often, and fail to use the 2 and 1 positions as often as intelligent, nonanxious *Ss*. The reverse effect is found for low IQ *Ss*. They appear to become more discriminating, utilizing the 1 and 2 positions more often when they are anxious than when they are not. The intelligent *S* seems to become more "all-or-nothing" in his responses when he is anxious, while the unintelligent *S* becomes more cautious or uncertain in his judgments.

The high IQ *Ss* in this experiment ranged from 106–127, a range within which most college students may be expected to fall. When we consider only the results for the high IQ quartile in the present experiment, the results are consistent with the studies previously cited (2, 6).

Osgood (4) suggests that there are three levels of difficulty in thinking in terms of gradients. When the means of judgment-times for all *Ss* using each position on the gradient were plotted, the obtained curve fitted the expected curve rather well. The shortest latency was found for the extreme positions (+3 and −3), a somewhat longer latency for the neutral position, and the longest latency for the intermediate positions (±2, ±1). In terms of frequency of usage, rather than latency of reaction-time, the present data support this hypothesis. The total group, as well as each individual group, used the extreme positions most, the neutral position somewhat less, and the intermediate positions least.

REFERENCES

1. BINDRA, D., PATERSON, A. L., & STRZELECKI, J. On the relation between anxiety and conditioning. *Canad. J. Psychol.*, 1955, **9**, 1–6.
2. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
3. KERRICK, J. S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75–77.
4. OSGOOD, C. E. Ease of individual judgment-processes in relation to polarization of attitudes in the culture. *J. soc. Psychol.*, 1941, **14**, 403–418.
5. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
6. SPENCE, K. W., & BEECROFT, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, **48**, 399–403.
7. TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Received January 24, 1955.

TAYLOR SCALE CHARACTERISTICS OF VOLUNTEERS AND NONVOLUNTEERS FOR PSYCHOLOGICAL EXPERIMENTS[1]

PHILIP HIMELSTEIN[2]

The University of Texas

A VAST body of psychological research has been based on *Ss* who have volunteered when approached in a classroom situation. Recently there has been some concern that personality differences between volunteers and nonvolunteers may introduce some bias into the experiment. Maslow and Sakoda (1), for example, found that volunteers for the Kinsey study on a college campus were higher in self-esteem than nonvolunteers, and that high self-esteem correlated with unconventional sex behavior. There is

[1] The author is indebted to Joseph S. Hillson, of Norfolk, Nebraska, State Hospital, for his helpful and constructive criticism in the preparation of this paper.

[2] Now at the Veterans Administration Hospital of Roanoke, Va.

also a suggestion from these writers that *S*s with extreme scores were volunteers rather than nonvolunteers.

The present study was designed to test the null hypothesis that there are no differences on the Taylor Manifest Anxiety Scale (MAS) between volunteers and nonvolunteers in a classroom setting.

Students in introductory classes in education and psychology, to whom *E* was unfamiliar, were employed for this experiment. The students were asked to indicate on a form their willingness or lack of willingness to serve as *S*s in a psychological experiment. The class was given no details concerning the experiment except that it would take approximately one-half hour, which would be arranged at the volunteering *S*'s convenience. Following the collection of these forms, the Taylor MAS was administered to the class as a whole.

A total of 141 *S*s were tested in this fashion, and 77, or 55 per cent of the total group, volunteered. The volunteer group obtained a mean MAS score of 14.25, and the nonvolunteer group had a mean score of 15.77. The difference (1.52 points) is not significant ($t = 1.23$; $.20 < p < .30$). A point biserial correlation of +.103 between the volunteer-nonvolunteer dichotomy and MAS score was obtained. These results would give no basis for rejecting the null hypothesis that there is no difference between the volunteer and nonvolunteer groups on the Taylor MAS.

Maslow and Sakoda have also suggested that *S*s with extreme scores on a test measuring self-esteem are more likely to volunteer than not volunteer. To test this hypothesis for the MAS, the scores were summarized in a 3×2 χ^2 table. The cut-off points for low and high scores were the 20th (score of 7) and 80th (score of 21) percentiles, respectively, of Taylor's (2) normative group. The resulting χ^2 of 2.67, with 2 *df*'s, is not significant ($p > .20$). This gives no reason to doubt that the two groups are alike on the distribution of extreme scores.

An incidental finding was that females, who constituted 54 per cent of the total sample, contributed 64 per cent to the volunteer group. Sex differences in volunteering might be further investigated with the sex and appearance of the individual who appeals for volunteers controlled.

It would appear that level of anxiety, to the extent that this factor is measured by the MAS, is not a major factor in the act of volunteering for psychological experiments. Apparently the use of *S*s obtained by appeals for volunteers does not introduce a bias into an experiment that is measurable by the Taylor MAS.

REFERENCES

1. MASLOW, A. H., & SAKODA, J. M. Volunteer-error in the Kinsey study. *J. abnorm. soc. Psychol.*, 1952, **47**, 259–262.
2. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Received January 4, 1955.

A METHOD FOR MATCHING ANONYMOUS QUESTIONNAIRE DATA WITH GROUP DISCUSSION MATERIAL

ELLIOTT McGINNIES

University of Maryland

IN view of the current interest of many social psychologists in discussion-group interaction and the number of research projects devoted to analysis of interaction variables, it may be helpful to investigators in this area to describe a technique for associating the comments made by discussion participants with biographical information obtained anonymously from them on questionnaires. These procedures have been devised as a means of securing data that will be used to describe the impact of mental health films upon community discussion groups.[1]

The aims of this research required that the following data be obtained: (*a*) verbatim recordings of group discussions following presentation of mental health films, (*b*) information concerning relevant socioeconomic and other characteristics of the individuals participating, (*c*) written reactions of group members to the film, so that the views of those who did not participate in the discussion could be evaluated. In order to relate this information, it was essential that some means be found of identifying the group members with their transcribed comments as well as with their questionnaires. At the same time, we recognized the importance of assuring anonymity to the participants. Any forthright attempt to identify individuals with either their spoken or written comments could have reduced seriously both the spontaneity

[1] This research was supported by Special Grant 3M9064 from the National Institute of Mental Health, United States Public Health Service. Mr. Irwin Altman, Mr. Willard Vaughan, and Mr. Clagett Smith assisted in developing the research procedures.

and the validity of their responses. It was also deemed necessary to record the discussions as inconspicuously as possible in order to further preserve the naturalness of the group sessions.

These various methodological problems were handled in the following ways. The groups to be studied met in their customary places. These ranged in size from school auditoriums to basement recreation rooms, depending upon the size and formal structure of the groups. Regardless of the physical setting, all groups were handled in substantially the same manner. Recording of the discussions was accomplished by placing from one to four microphones at strategic positions in the conference rooms. All microphones were of the semi-directional type. They fed into a standard commercial mixer, which made it possible to record from them individually through a twin-track tape recorder operating at 7½ inches per second. To avoid cross-talk interference, only the microphone nearest a discussion participant was opened at any one time. All recording apparatus with the exception of the microphones was located in the rear of the room where its presence was not obtrusive. The group, of course, had been informed that the discussion was to be recorded.

Printed questionnaires were distributed at the beginning of the meeting, and the group members were asked to supply the indicated personal background information. They were told not to sign their names, since the answers were to remain anonymous. At the conclusion of the film they were asked to respond to an open-end question relating to film content. Finally, following a half-hour discussion of the film, the group members answered two questions pertaining to the discussion. They were instructed to leave their questionnaires upon their seats at the conclusion of the meeting. Identification of the questionnaires of individuals who spoke during the discussion, without violation of our promise of anonymity, was accomplished readily through some advance preparation by the experimenters. From a preliminary examination of the meeting place, we prepared a chart indicating each row by letter and each seat by number. Audience members were permitted to sit anywhere. When someone arose to speak during the discussion, his location by row and seat was recorded by two observers stationed at the rear of the room. An exact, seriatim account of the location of each participant was thus obtained throughout the discussion period. At the conclusion of the meeting, the questionnaires were collected in order, and each was identified by the row and seat of its owner.

In our laboratory, the taped recording of the discussion was transferred to an Edison Voicewriter disc for transcribing by a secretary. By listening to the tape with the transcription before us, we were able to identify each person who spoke from the records kept by the two spotters during the discussion. The first speaker, for example, may have occupied seat B4, the second speaker seat G6, etc. The first transcribed comment, therefore, was that of B4, the second comment that of G6. Thus, without knowing the names of any of the discussion participants, we were able to assign a symbol to each of them and to designate both their verbal comments and their questionnaires by that symbol. The fact that two observers recorded the order of participation enabled us to check the reliability of the procedure. Agreement between the spotters has consistently been around 98 per cent.

We hope that this methodological note will be helpful to those who are studying group interaction and who may require the type of information described here. From data such as we have acquired it is possible to compare individuals who participate in group discussion with those who do not participate. Sociometric data obtained from questionnaires can be used as predictors of individual participation and interaction in the discussion situation. Perhaps the main value of this approach is that of being able to study such relationships among persons who might be reluctant to participate if their names were used for purposes of identification.

Received January 7, 1955.

THE FAKABILITY OF TWO MEASURES OF PREJUDICE: THE CALIFORNIA F SCALE AND GOUGH'S *Pr* SCALE

NORMAN D. SUNDBERG AND WARREN D. BACHELIS
University of Oregon *Harvard University*

ONE of the most widely used measures in studies of prejudice is the F (Fascism) scale developed in the University of California investigation of the authoritarian personality. This scale has been an important subject in two major books (1, 2) and a large number of articles. Another measure of prejudice is *Pr* (Prejudice), a scale developed by Gough (4) from the MMPI item pool and now used with reverse scoring as the Tolerance scale in his new California Psychological Inventory.

There has been little concern with the ability of subjects to fake a prejudiced or unprejudiced picture of themselves on these two measures. One exception is an article by Cohn (3), in which it was reported that a sample of college students could successfully fake an unprejudiced score when asked to answer the F scale as highly educated, intelligent people would. They could also successfully fake a prejudiced score when asked to answer as unintelligent laborers would. Cohn reported a positive relationship between ability

to fake and intelligence test scores. He concluded that the F scale is not the indirect test it purports to be. Hollander (5) found that naval cadets shifted their responses significantly when given the set to answer the F scale as if it were a leadership test on which they wished to make a good score. In some studies of shift of scores on the F scale accompanying training, it may well be that the change is due to an increase in sophistication, not a basic personality change. There is no report in the literature on the susceptibility of the *Pr* scale to faking. Since these scales are used in many studies and since test-taking attitude is known to affect scores on many tests, it seems important to have evidence as to the possible effect of intent to deceive on these scales.

PROCEDURE

The subjects (*Ss*) were students in an elementary psychology course. Form 60 of the F scale was used. All *Ss* who reported having seen the items on the F scale beforehand were excluded from the final tabulations. A few who belonged to so-called minority groups were excluded from the analysis. These were *Ss* not reporting white race, Protestant or Catholic religion, and native American birth on questions asked at the end of the testing period. These exclusions left a total of 52.

The *Ss* were not required to put identifying data on their papers other than sex and date of birth. (These data were used in order to match up the different sets of the scales.) After the *Ss* had answered the items on both tests in the regular manner, they were given another blank set of the scales and the following instructions were read to them:

> Psychologists have long been interested in the ability of people to respond to questions as they think another person would answer them. This ability "to put yourself into someone else's shoes," so to speak, is called empathy. You could have been asked, for example, to answer the two questionnaires you just took as you thought a particular friend might have answered them, and so on. I'm going to ask you to answer the same two questionnaires again but this time in the way you think they would be answered by a person who is generally prejudiced and intolerant. Each of you probably has some notion of what the prejudiced person is like. Will you then answer the same set of questions as you think it would be answered by a very prejudiced person?

When these forms were completed, *Ss* were given another set of the scales and they were instructed as follows:

> This time you will be asked to answer the questionnaires in the way that you think they would be answered by a person who is tolerant and unprejudiced. I won't attempt to define tolerance or lack of prejudice. Each of you probably has some notion as to what the tolerant and unprejudiced person is like.

RESULTS AND DISCUSSION

The data gathered from the *Ss* consisted of three scores for each of the two prejudice scales: a regular score under conventional conditions of administration, a fake-prejudiced score and a fake-unprejudiced score. Table 1 presents the means and standard deviations for both scales for each of the three conditions.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SCORES OF THE 52 *Ss* ON THE F SCALE AND *Pr* SCALE UNDER THE THREE TEST-TAKING CONDITIONS

	Regular	Prejudiced	Unprejudiced
F Scale*			
Mean	102.70	185.00	100.98
SD	24.38	21.79	27.63
Pr Scale			
Mean	6.58	18.55	3.06
SD	4.17	5.60	2.48

* The score for the F scale is the total of the scale values for a given S (1, p. 72). Since there are 33 items on Form 60 of the F scale, the "most unprejudiced" score would be 33; the "most unprejudiced" score, 231. On the *Pr* scale, the range would be from 0 on the "unprejudiced" end to 32 on the "prejudiced" end.

On both scales the subjects produced scores significantly more prejudiced than their regular scores. The difference between the means of F-Regular and F-Prejudiced scores and the difference between the means of *Pr*-Regular and *Pr*-Prejudiced scores were both significant at the .01 level of confidence. Overlap between the distributions for the regular and fake-prejudiced scores was slight on both scales. Except for one *S* on the *Pr* scale, all *Ss* answering the F or *Pr* scales in the regular manner scored below the lowest quartile of the corresponding fake-prejudiced distributions.

The results of the attempt to fake an unprejudiced score were not as clear as the faking of the prejudiced attitude. With both scales less prejudiced scores were obtained, but the mean differences between the regular and unprejudiced distributions was significant only for the *Pr* scale.

It would have been possible for subjects to lower their F scores considerably, but they did not. Among plausible reasons for this outcome are the following: (*a*) subtlety of the F scale with regard to faking in the unprejudiced direction, (*b*) the possibility that very low F scores may not signify lack of prejudice, and (*c*) the possibility that the *Ss*' regular scores were already faked in the unprejudiced direction. Our group's average regular scores were already lower than those of the groups reported in the *Authoritarian Personality* (1, p. 266) and it would seem that they started near the bottom on the continuum of prejudice scores, and thus would have difficulty shifting still lower.

The product-moment correlation coefficient between the regular F and *Pr* scores was +.35, a figure significantly different from 0 at the one per cent level of confidence. This coefficient is

somewhat lower than the figures reported by Gough (4).

The data offer an opportunity to answer the question, "Is ability to fake the prejudiced attitude related to the *S*'s own degree of prejudice?" Using the regular score as a measure of prejudice, we found the correlation between it and the fake-prejudiced score to be −.16 on the F scale and .20 on the *Pr* scale. Neither of these figures is significantly different from a correlation of zero.

SUMMARY

Fifty-two elementary psychology students, who were naive to the tests and who did not belong to minority groups, were given the California F (Authoritarian) scale and Gough's *Pr* (Prejudice) scale. First, they took the scales with the regular instructions. They repeated them, once faking prejudiced attitudes, and again faking unprejudiced attitudes. The subjects were clearly able to fake the prejudiced approach. However, they were able to change their scores significantly toward the unprejudiced direction only on the *Pr* scale. The F and *Pr* regular scores correlated .35. No relationship was found between regular prejudice scores and ability to fake prejudice. The results imply that both the F and *Pr* scales are subject to the influence of test-taking attitudes. It appears particularly easy to fake prejudice if the person is so inclined.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. CHRISTIE, R., & JAHODA, MARIE (Eds.). *Studies in the scope and method of "The authoritarian personality"*. Glencoe, Ill.: Free Press, 1954.
3. COHN, T. S. Is the F scale indirect? *J. abnorm. soc. Psychol.*, 1952, **47**, 732.
4. GOUGH, H. G. Studies of social intolerance, I, II, III, and IV. *J. soc. Psychol.*, 1951, **33**, 237–246, 247–255, 257–262, 263–269.
5. HOLLANDER, E. P. Authoritarianism and leadership choice in a military setting. *J. abnorm. soc. Psychol.*, 1954, **49**, 365–370.

Received January 24, 1955.

The International Journal of Social Psychiatry

THE FIRST INTERNATIONAL QUARTERLY IN THE FIELD OF INTERDISCIPLINARY RESEARCH

Winter Issue

The Third Issue will include articles by:—

LEONARD J. DUHL. Psychiatry and the Community.
ERIKA CHANCE AND STUART E. ATKINSON. Some Interpersonal Characteristics of Individual Treatment.
RUDOLF DREIKURS. Group Psychotherapy and the Third Revolution in Psychiatry.
G. M. CARSTAIRS. Attitudes to Death and Suicide in an Indian Cultural Setting.
MOUSTAFA TOLBA. Psychiatric Observations of International Conferences.
H. A. D. BUCHMULLER AND MARGARET C.-L. GILDEA. Group Therapy for Parents of Behaviour Problem Children in Public Schools.
NUÑEZ MOLINA. Psychiatric Conditions in Spain.
C. M. FLEMING. Work in Progress.
Reports, Reviews and Correspondence.

SUBSCRIPTIONS:

Great Britain and Sterling Area.	Germany.	U. S. A., Canada, and other countries
1 or Issue. 7s. 6d.	6DM.	$1.50
Four Issues. 25s. 0d.	20DM.	$5.

All enquiries should be addressed to:

The Secretary, International Journal of Social Psychiatry,
9 Fellows Road London, N.W. 3. England.

INTERNATIONAL SOCIAL SCIENCE BULLETIN

published quarterly by the United Nations Educational, Scientific and Cultural Organization 19, Avenue Kléber, Paris 16e

JUST ISSUED: VOL. VII, NO. 3

EVALUATION TECHNIQUES

METHODS AND RESULTS—AN INTERNATIONAL APPROACH

Contributors to this issue: E. Barnitz, E. Beaglehole, M. Jahoda, O. Klineberg, W. Maslow, K. M. Miller, L. Moss, C. Selltiz, M. Brewster Smith, C. R. Wright.

ORGANIZATION IN THE SOCIAL SCIENCES—TERMINOLOGY—REVIEWS OF DOCUMENTS, PERIODICALS AND BOOKS—NEWS AND ANNOUNCEMENTS

OPEN FORUM:

The Uses of the Law in the Struggle Against Bigotry, by Wil Maslow

Send subscriptions to:
UNESCO PUBLICATIONS CENTER
475 Fifth Avenue, New York 17, N.Y.

Annual subscription: $3.50
Per copy: $1.00

Coming in February

What We Learn From Children

By MARIE L. RASEY and J. W. MINGE
Professors of Education, Wayne University

Two wise and successful teachers, drawing on fruitful experience in teaching and nurturing problem children, here re-examine our assumptions about how children learn and grow, and offer new perspectives on helping children, based on a child's own needs and ways of reacting to experience. Vividly illustrated with stories of children who were helped to overcome a wide variety of difficulties.

"An outstanding example of fundamental research in education . . . an example of knowledge culminating in wisdom."—ROSS L. MOONEY, *Professor of Education, Ohio State University*

$3.00
at your bookstore or from
HARPER
&
BROTHERS
49 E. 33rd St.
N. Y. 16

BORZOI College Textbooks

Case Book in Abnormal Psychology

by HENRY WEINBERG *and* A. WILLIAM HIRE
Boston University

A collection of twenty individual cases which cover the major areas of psychological abnormality, these studies emphasize not merely the period of breakdown and the psychological test results, but the function and development of the disorder in the total growth of the individual. Since the cases are presented without interpretation and with no comment beyond a general introduction to the book, this text can be used for any theoretical orientation and for whatever depth of analysis is desired.

Ready in February, 320 pages, $4.50

ALFRED A. KNOPF, *Publisher*
501 Madison Ave. *College Department* New York